Journal of Multivariate Analysis 103 (2012) 124–141

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Nonparametric and semiparametric optimal transformations of markers

Chin-Tsang Chiang*, Chih-Heng Chiu

Department of Mathematics, National Taiwan University, Taipei 10617, Taiwan, ROC

ARTICLE INFO

Article history:
Received 15 December 2010
Available online 30 June 2011

AMS subject classifications:
62G07
62G08
62G20
62H30

Keywords:
Area under curve
Nonparametric estimator
Optimal marker
Receiver operating characteristic curve
Single-index model
U-statistic

ABSTRACT

The receiver operating characteristic (ROC) curve of a likelihood-ratio function has been shown to be the highest among all transformations of continuous markers. For any sampling scheme with the same likelihoods, the induced conditional probability is derived to have the same ROC curve and is found to be more useful for inference purposes. To compromise the difficult task of high-dimensionality in fully nonparametric models and the risk of model misspecification in fully parametric ones, an appealing single-index model is also adopted in our optimization problem. Based on a nonparametric estimator of the area under the ROC curve (AUC), we develop its related inferences and provide some simple and easily checked conditions for the validity of asymptotic results. Since the optimal marker is estimated by using a semiparametric or nonparametric model, conventional theoretical approaches might be inappropriate to some circumstances. The applicability of our procedures are further demonstrated through extensive numerical experiments and data from the studies of Pima–Indian diabetes and liver disorders.

1. Introduction

This article deals with a two-category classification problem of continuous markers which is frequently encountered in clinical research, signal detection practice, biological science, and many other fields. Let $D \in \{0, 1\}$, $Y = (Y_1, \ldots, Y_q)^\top$, and $\mathcal{C}_u(Y) \in \{0, 1\}$ denote the true class, the continuous marker(s), and a classifier of Y, respectively, with $f_d(y)$ being the conditional density function of Y given $D = d$ and $P(\mathcal{C}_u(Y) = 1|D = 0) = u$. The true positive rate and false positive rate are naturally defined by $\mathcal{T}_{C_u} = P(\mathcal{C}_u(Y) = 1|D = 1)$ and $\mathcal{F}_{C_u} = P(\mathcal{C}_u(Y) = 1|D = 0)$. Attention usually focuses on seeking the best classifier $\mathcal{C}_u^{opt}(Y)$ on which $\mathcal{T}_{C_u^{opt}} \geq \mathcal{T}_{C_u}$. Applying the argument of [15] for the hypotheses $H_0 : \{D = 0\}$ versus $H_A : \{D = 1\}$ with the significance level of u, a classifier based on the likelihood-ratio function $\ell(Y) = f_1(Y)/f_0(Y)$ can be shown to be $\mathcal{C}_u^{opt}(Y)$. The spectrum of $(\mathcal{F}_{C_u^{opt}}, \mathcal{T}_{C_u^{opt}})$ over $u \in [0, 1]$ forms the ROC curve of $\ell(Y)$ and is often used to assess its overall capability to correctly classify subjects as $\{D = 0\}$ or $\{D = 1\}$.

For any marginal probability p_d of $\{D = d\}$, $g(y) = p_1\ell(y)/(p_0+p_1\ell(y))$ is shown to be a strictly increasing transformation of $\ell(y)$. This property implies the optimality of $g(Y)$ and a direct algebraic computation yields that

$$g(y) = \frac{p_1 f_1(y)}{p_0 f_0(y) + p_1 f_1(y)} \quad \text{with } g(y) = P(D = 1|Y = y) \text{ whenever } p_1 = P(D = 1). \tag{1.1}$$

Interestingly, $g(y)$ can be explained as the probability under a new sampling scheme (cf. [19]) and is very useful to accommodate different study designs from a practical point of view. As demonstrated in the literature, the accuracy measure AUC is superior in comparing the discriminating abilities of continuous markers when their ROC curves are identical or

* Corresponding author.
E-mail address: chiangct@ntu.edu.tw (C.-T. Chiang).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2011.06.014

non-crossed. Since the existing inferences might be problematic under some circumstances, we develop a more complete inference for the AUC of optimal markers. In the case of $q = 1$ (a univariate marker), a frequent issue confronting the general practitioner is to detect whether the marker itself is a potential optimal marker. Currently, there exists no rigorous testing procedure to this end. Following the empirical standard of [10], the discriminating ability of individual markers might not be acceptable in some applications. An important research issue and challenge is to seek an optimal composite marker of multiple markers. The primary aim of assessment is to screen out unimportant markers and identify a correct subset $Y_s = (Y_{s_1}, \ldots, Y_{s_m})^\top$, where $s = (s_1, \ldots, s_m)$ and $1 \le m \le q$, of Y with its scoring function $g_m^q(Y_s)$ and $g(Y) = g_q^q(Y_s)$ having the same or close classification accuracy.

Another scenario for this optimization task is to explore whether $g_m^q(y_s)$ is a function of a specific linear combination of markers of the form $y_{\beta_s} = \beta_s^\top y_s$ with $\beta_s = (\beta_{s_1}, \ldots, \beta_{s_m})^\top$. In applications, generalized linear models have played prominent roles and have been widely used to characterize the influence of markers on a binary outcome probability. Examples of such parametric models include the logistic, the complementary log–log, and the probit regressions, among others. The related inference procedures were comprehensively studied in the literature. To compromise the risk of misspecified parametric models and the poor results in fully nonparametric ones, we consider and investigate a more acceptable single-index model (SIM):

$$P(D = 1|Y_s = y_s) \stackrel{\Delta}{=} P(D = 1|Y_{\beta_s} = y_{\beta_s}), \tag{1.2}$$

where $P(D = 1|Y_{\beta_s} = y_{\beta_s})$ is an unspecified function of y_{β_s}. It is easily derived from (1.1)–(1.2) that $g_m^q(y_s)$ is also a function of y_{β_s}, say, $g_{msi}^q(y_{\beta_s})$. The pseudo maximum likelihood estimation approach of [12] and the pseudo least squares one of [11] can be adopted to estimate the scoring function $g_{msi}^q(y_{\beta_s})$. Under the monotonic assumption on $g_{msi}^q(y_{\beta_s})$, [13,17] developed the related inference procedures via incorporating an estimator of Y_{β_s} into the empirical and sigmoid AUC objective functions, respectively. As one can see from our examples, some degree of modeling of y_{β_s} might be inappropriate and will lead to a non-optimal scoring function. A more flexible model should be of great interest to characterize the dependence of D on Y_{β_s}.

In the next section, we propose estimators for the AUCs of $g_m^q(Y_s)$ and $g_{msi}^q(Y_{\beta_s})$. The asymptotic properties and inference procedures are developed in Section 3. In Section 4, simulation studies are carried out for performance evaluation. The following section covers applications of our approaches to two empirical examples. Some discussion and research issues are provided in Section 6 and the proofs of the main results are placed in the Appendix.

2. Estimation and bandwidth selection

In this section, we consider a random sample of the form $\{(D_i, Y_i)\}_{i=1}^n$, which might be obtained from either a prospective study or a retrospective study (e.g. case-control design). For ease of presentation and better illustration, we let $n_1 = \sum_{i=1}^n D_i$, $n_0 = n - n_1$, and θ_M denote the AUC of marker M with or without of transformation. Details of estimation procedures for $\{g_m^q(Y_s), \theta_{g_m^q}\}$ and $\{g_{msi}^q(Y_{\beta_s}), \theta_{g_{msi}^q}\}$ are presented below.

With data from a prospective study, $g_m^q(y_s)$ is naturally estimated by the Nadaraya–Watson type kernel estimator:

$$\widehat{g}_m^q(y_s) = \frac{\sum_{i=1}^n \prod_{j=1}^m K_{h_{1j}}(y_{s_j} - Y_{is_j})D_i}{\sum_{i=1}^n \prod_{j=1}^m K_{h_j}(y_{s_j} - Y_{is_j})}, \tag{2.1}$$

where $K_\varsigma(u) = K(u/\varsigma)/\varsigma$ and $K(\cdot)$ is a density function. To overcome the difficulty occurring in a high-dimensional bandwidth space, the bandwidths $h_{1j} = s_{1j}h$ and $h_j = s_j h$ are specified in our numerical studies, where h is a positive-valued smoother, s_j is the standard deviation of $\{Y_{is_j}\}_{i=1}^n$, and s_{1j} is the standard deviation of Y_{is_j}'s with D_i's being one, $j = 1, \ldots, m$. In a retrospective study $p_1 = \lim_{n_0, n_1 \to \infty} n_1/(n_0 + n_1)$, the estimator in (2.1) can also be used in the estimation of $\theta_{g_m^q}$. For $q = 1, g(y)$ can also be estimated by

$$\widehat{g}(y) = \frac{\widehat{p}_1\widehat{f}_1(y)}{\widehat{p}_0\widehat{f}_0(y) + \widehat{p}_1\widehat{f}_1(y)} = \frac{\sum_{i=1}^n K_{h_1}(y - Y_i)D_i}{\sum_{i=1}^n K_{h_0}(y - Y_i)(1 - D_i) + \sum_{i=1}^n K_{h_1}(y - Y_i)D_i} \tag{2.2}$$

with $\widehat{p}_d = n_d/n$ and h_d being a positive-valued bandwidth, $d = 0, 1$. Since the scoring function $g_m^q(y_s)$ is usually unknown in practice, $\theta_{g_m^q}$ is proposed to be estimated by the following Mann–Whitney U-statistic with the correction of ties:

$$\widehat{\theta}_{\widehat{g}_m^q} = (n_0 n_1)^{-1} \sum_{i \ne j} [I\{\widehat{g}_m^q(Y_{is}) > \widehat{g}_m^q(Y_{js})\} + 0.5I\{\widehat{g}_m^q(Y_{is}) = \widehat{g}_m^q(Y_{js})\}]D_{ij}, \tag{2.3}$$

where $D_{ij} = D_i(1 - D_j)$. Although $\widehat{\theta}_{\widehat{g}_m^q}$ still generally suffers from the curse of dimensionality [1], this problem is not so serious as in $\widehat{g}_m^q(y_s)$ because the accuracy measure strongly depends on the ranks of $\widehat{g}_m^q(Y_{is})$'s. In the next section, both of $\widehat{\theta}_{\widehat{g}_m^q}$

and $\widehat{\theta}_{g_m^q}$, which is computed as $\widehat{\theta}_{\widehat{g}_m^q}$ with $g_m^q(Y_{is})$'s being substituted for $\widehat{g}_m^q(Y_{is})$'s, will be shown to have the same asymptotic distribution under some suitable and easily checked conditions.

By extending the estimation methods of [11,12] to accommodate any sampling design, the scoring function $g_{msi}^q(y_{\beta_s})$ is proposed to be estimated by

$$\widehat{g}_{msi}^q(y_{\widehat{\beta}_s}) = \frac{\sum_{i=1}^n K_h(Y_{i\beta_s} - y_{\beta_s})D_i}{\sum_{i=1}^n K_h(Y_{i\beta_s} - y_{\beta_s})} \quad \text{with } \widehat{\beta}_s = \arg\min_{\beta_s} \sum_{i=1}^n \{D_i - \widehat{g}_{msi}^q(y_{\beta_s})\}^2. \tag{2.4}$$

For the sake of identifiability in the SIM, the widely used constraint $\beta_s^\top \beta_s = 1$ is adopted. The methods in [4] can also be considered as an alternative to estimate the index-coefficients, while a stronger assumption on the distribution of markers is usually required to ensure the consistency of their estimators. By applying the Mann–Whitney statistic as in (2.3), we can directly obtain an estimator of $\theta_{g_{msi}^q}$ as

$$\widehat{\theta}_{\widehat{g}_{msi}^q} = (n_0 n_1)^{-1} \sum_{i \neq j} [I\{\widehat{g}_{msi}^q(Y_{i\widehat{\beta}_s}) > \widehat{g}_{msi}^q(Y_{j\widehat{\beta}_s})\} + 0.5 I\{\widehat{g}_{msi}^q(Y_{i\widehat{\beta}_s}) = \widehat{g}_{msi}^q(Y_{j\widehat{\beta}_s})\}] D_{ij}. \tag{2.5}$$

For the bandwidths in $\widehat{g}_m^q(y_s)$ and $\widehat{g}_{msi}^q(y_{\widehat{\beta}_s})$, the smoothers h_{1cv} and h_{2cv} are separately selected to be the minimizers of

$$\sum_{i=1}^n \{D_i - \widehat{g}_{m(-i)}^q(Y_{is})\}^2 \quad \text{and} \quad \sum_{i=1}^n \{D_i - \widehat{g}_{msi(-i)}^q(Y_{i\widehat{\beta}_{s(-i)}})\}^2, \tag{2.6}$$

where $\widehat{g}_{m(-i)}^q(y_s)$ and $\widehat{g}_{msi(-i)}^q(y_{\beta_s})$, $\widehat{\beta}_{s(-i)}$ are computed as their counterparts with the ith individual being deleted. Another criterion developed by [7] for $g_m^q(y_{\beta_s})$ is to simultaneously estimate β_s and h by minimizing

$$S(\beta_s, h) = \sum_{i=1}^n \{D_i - \widehat{g}_{msi(-i)}^q(Y_{i\beta_s})\}^2. \tag{2.7}$$

As for the smoothing estimator in (2.2), the bandwidth selection procedure of [20] can also be applied to $\widehat{f}_0(y)$ and $\widehat{f}_1(y)$. Different from the conclusion drawn from [21] for $\widehat{g}(y)$, $\widehat{\theta}_{\widehat{g}}$ with the two-separate-bandwidth estimator in (2.2) outperforms those with the single-bandwidth and two-bandwidth estimators in (2.1) (not shown here).

3. Inference procedures

The asymptotic equivalence between $\widehat{\theta}_{\widehat{g}_m^q}$ and $\widehat{\theta}_{g_m^q}$ is first introduced in this section. Both of $\widehat{\theta}_{\widehat{g}_{msi}^q}$ and $\widehat{\theta}_{g_{msi}^q}$ are also shown to be asymptotically equivalent in a similar manner. Moreover, we construct various confidence intervals for $\theta_{g_m^q}$ and $\theta_{g_{msi}^q}$, and establish some testing procedures for the optimality of a proposed marker in classification.

3.1. Asymptotic equivalence

Since $g_m^q(y_s)$ is a function of any bounded and strictly increasing transformation of y_s, the support $\mathcal{Y}_s$ of Y_s is reasonably assumed to be compact in theoretical developments. Let $f(y_s) = p_0 f_0(y_s) + p_1 f_1(y_s)$ with $m_f = \inf_{y_s} f(y_s)$ and $M_f = \sup_{y_s} f(y_s)$, $f_m(y_{1s}, y_{2s}) = \min\{f(y_{1s}), f(y_{2s})\}$, and

$$I(\varepsilon) = \sup_{0 \le u \le M_f} \int_{\{u \le f_m(y_1, y_2)\{g_m^q(y_{1s}) - g_m^q(y_{2s})\}^2 \le u + \varepsilon\}} I\{g_m^q(y_{1s}) \neq g_m^q(y_{2s})\}\{f(y_{1s}) f(y_{2s})\}^{1/2} \mathrm{d}y_{1s} \mathrm{d}y_{2s},$$

which is a nondecreasing and continuous function of $\varepsilon \ge 0$. Some assumptions are further made below.

(A1) $h \to 0$, $n \to \infty$, and $nh^m \to \infty$.
(A2) $\sup_{y_s} f_d(y_s) < \infty$.
(A3) $g_m^q(y_s)$ is twice continuously differentiable with bounded derivatives.
(A4) $I((nh)^{-1}) = o(h^{1/2})$, $\int \mathrm{e}^{-nhf_m(y_{1s}, y_{2s})} f(y_{1s}) f(y_{2s}) \mathrm{d}y_{1s} \mathrm{d}y_{2s} = o(n^{-1/2})$, $\int_{\{g_m^q(y_{1s}) = g_m^q(y_{2s})\}} f(y_{1s}) f(y_{2s}) \mathrm{d}y_{1s} \mathrm{d}y_{2s} = 0$.

From assumption (A2), one can easily demonstrate that $M_f < \infty$. When $m_f > 0$, the second condition in assumption (A4) automatically holds and $I(\varepsilon)$ will be simplified to

$$I_0(\varepsilon) = \sup_{0 \le u \le M_f} \int_{\{u \le \{g_m^q(y_{1s}) - g_m^q(y_{2s})\}^2 \le u + \varepsilon\}} I\{g_m^q(y_{1s}) \neq g_m^q(y_{2s})\} \mathrm{d}y_{1s} \mathrm{d}y_{2s}.$$

For a univariate marker, $dg(y)/dy > 0$ (or $dg(y)/dy < 0$) and $\inf_y |dg(y)/dy| > 0$ implies

$$I_0((nh)^{-1}) \le \sup_{0\le u\le 1} \int_{\{u^{1/2}\le |g(y_1)-g(y_2)|\le u^{1/2}+(nh)^{-1/2}\}} I\{g(y_1)\ne g(y_2)\}dy_1dy_2$$
$$\le 2(\inf_y |dg(y)/dy|)^{-1}(nh)^{-1/2} = o(h^{1/2}) \quad \text{for some } h.$$

If $g(y)$ has finite critical points and locally behaves like a power function in a small neighborhood of each critical point with the maximum degree $\nu > 0$, it can be shown that $h^{-1/2}I_0((nh)^{-1}) = O((nh^{1+\nu})^{-1/2\nu}) = o(1)$ for $nh^{1+\nu} \to \infty$. Our conjecture is that there exists a bandwidth h satisfying assumption (A4) and the first condition in (A4) whenever $I(0) = 0$ and $I(M_f) < \infty$. Note that the second condition in (A4) is mainly made for the shape of $f(y_s)$ and a heavy-tailed distribution such as the Cauchy distribution would cause its violation. When $g_m^q(y_s)$ is flat in some regions, the third condition in (A4) is automatically violated and the asymptotic properties of $\widehat{\theta}_{g_m^q}$ cannot apply to $\widehat{\theta}_{\widehat{g}_m^q}$. The first theorem is given by the following:

Theorem 1. *Under assumptions* (A1)–(A4), $\sqrt{n}(\widehat{\theta}_{\widehat{g}_m^q} - \widehat{\theta}_{g_m^q}) \xrightarrow{p} 0$ *as* $n \to \infty$.

Proof. See the Appendix. □

For the asymptotic equivalence between $\widehat{\theta}_{\widehat{g}_{msi}^q}$ and $\widehat{\theta}_{g_{msi}^q}$, some basic notations are introduced below. Let $\mathcal{B}_s$ be the support of β_s, $f_{sd}(y_{\beta_s})$ denote the conditional density of Y_{β_s} given $D = d$ with support $\mathcal{Y}_{\beta_s} = \{y_{\beta_s} : y_s \in \mathcal{Y}_s, \beta_s \in \mathcal{B}_s\}$, and $f_s(y_{\beta_s}) = p_0 f_{s0}(y_{\beta_s}) + p_1 f_{s1}(y_{\beta_s})$ with $f_{sm}(y_{1\beta_s}, y_{2\beta_s}) = \min\{f_s(y_{1\beta_s}), f_s(y_{2\beta_s})\}$, $m_{f_s} = \inf_{y_{\beta_s}} f_s(y_{\beta_s})$, and $M_{f_s} = \sup_{y_{\beta_s}} f_s(y_{\beta_s})$. In addition, we define

$$I_{\beta_s}(\varepsilon) = \sup_{0\le u\le M_{f_s}} \int_{\{u\le\{g_{msi}^q(y_{1\beta_s})-g_{msi}^q(y_{2\beta_s})\}^2\le u+\varepsilon\}} I\{g_{msi}^q(y_{1\beta_s})\} \ne g_{msi}^q(y_{2\beta_s}))dy_{1\beta_s}dy_{2\beta_s}$$

and impose the following regularity conditions:

(B1) $\mathcal{B}_s$ is compact and β_s is an interior point of $\mathcal{B}_s$.
(B2) $K(u)$ is symmetric and twice continuously differentiable in $[-1, 1]$ with its second derivative satisfying a Lipschitz condition.
(B3) $h \to 0, n \to \infty, nh^5 \to \infty$, and $nh^8 \to 0$.
(B4) $\inf_{y_{\beta_s}} f_{sd}(y_{\beta_s}) > 0$ and $\sup_{y_{\beta_s}} f_{sd}(y_{\beta_s}) < \infty$.
(B5) $f_s(y_{\beta_s})$ and $E[Y_s|Y_{\beta_s} = y_{\beta_s}]$ are three times continuously differentiable with bounded derivatives.
(B6) $g_{msi}^q(y_{\beta_s})$ is twice continuously differentiable with bounded derivatives.
(B7) $I_{\beta_s}((nh)^{-1}) = o(h^{1/2})$, $\int e^{-nhf_{sm}(y_{1\beta_s}, y_{2\beta_s})} f_s(y_{1\beta_s}) f_s(y_{2\beta_s}) dy_{1\beta_s} dy_{2\beta_s} = o(n^{-1/2})$, and $\int_{\{g_{msi}^q(y_{1\beta_s})=g_{msi}^q(y_{2\beta_s})\}} f_s(y_{1\beta_s}) f_s(y_{2\beta_s}) dy_{1\beta_s} dy_{2\beta_s} = 0$.

As those in [11], assumptions (B1)–(B5) are standard conditions in semiparametric inferences and are often used for the $\sqrt{n}$-consistency and the asymptotic normality of $\widehat{\beta}_s$. Here, $\inf_{y_{\beta_s}} f_{sd}(y_{\beta_s}) > 0$ enables $I_{\beta_s}(\varepsilon)$ to have the same explanation of $I_0(\varepsilon)$ and assumptions (B6)–(B7) are slight modifications of assumptions (A3)–(A4). The main result of $\widehat{\theta}_{\widehat{g}_{msi}^q}$ is stated below.

Theorem 2. *Under assumptions* (B1)–(B7), $\sqrt{n}(\widehat{\theta}_{\widehat{g}_{msi}^q} - \widehat{\theta}_{g_{msi}^q}) \xrightarrow{p} 0$ *as* $n \to \infty$.

Proof. See the Appendix. □

3.2. Confidence intervals

For ease of exposition, we let θ_g denote $\theta_{g_m^q}$ or $\theta_{g_{msi}^q}$ and η_{ij} represent $I\{g_m^q(Y_{is}) > g_m^q(Y_{js})\}$ or $I\{g_{msi}^q(Y_{i\beta_s}) > g_{msi}^q(Y_{j\beta_s})\}$. By substituting $(\widehat{g}_m^q(Y_s), \widehat{g}_{msi}^q(Y_{\widehat{\beta}_s}))$ for $(g_m^q(Y_s), g_{msi}^q(Y_{\beta_s}))$, $\widehat{\theta}_{\widehat{g}}$ and $\widehat{\eta}_{ij}$ are directly obtained. Since $\widehat{\theta}_g$ is a U-statistic, one has

$$\sqrt{n}(\widehat{\theta}_g - \theta_g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Gamma_{gi} + o_p(1) \quad \text{(cf. [9])}, \tag{3.1}$$

where $\Gamma_{gi} = p_1^{-1}E[\eta_{ij}D_{ij} - \theta_g|Y_{is}, D_i] + p_0^{-1}E[\eta_{ji}D_{ji} - \theta_g|Y_{is}, D_i]$, $i = 1, \ldots, n$. The following theorem is ascertained through the decomposition $\sqrt{n}(\widehat{\theta}_{\widehat{g}} - \theta_g) = \sqrt{n}(\widehat{\theta}_{\widehat{g}} - \widehat{\theta}_g) + \sqrt{n}(\widehat{\theta}_g - \theta_g)$, the limiting distribution of $\sqrt{n}(\widehat{\theta}_g - \theta_g)$, and Theorems 1–2.

Theorem 3. *Suppose that the assumptions in Theorem 1 or Theorem 2 are satisfied. Then,*

$$\frac{\sqrt{n}(\widehat{\theta}_{\widehat{g}} - \theta_g)}{\sigma_g} \xrightarrow{d} N(0, 1) \quad \text{as } n \to \infty, \text{ where } \sigma_g^2 = E[\Gamma_{g1}^2]. \tag{3.2}$$

From the asymptotic normality of $\widehat{\theta}_{\widehat{g}}$, an approximated $100(1-\alpha)\%$ confidence interval can be constructed for θ_g for any $\alpha \in (0, 1)$. It is given by

$$\widehat{\theta}_{\widehat{g}} \pm z_{1-\alpha/2}\, \mathrm{se}(\widehat{\theta}_{\widehat{g}}), \tag{3.3}$$

where $\mathrm{se}(\widehat{\theta}_{\widehat{g}}) = n^{-1/2}\widehat{\sigma}_g$ with $\widehat{\sigma}_g^2 = n^{-1}\sum_{i=1}^n \widehat{\Gamma}_{\widehat{g}i}^2$ being a consistent estimator of σ_g^2 and $\widehat{\Gamma}_{\widehat{g}i} = n(n_0n_1)^{-1}\sum_{j=1}^n(\widehat{\eta}_{ij}D_{ij} + \widehat{\eta}_{ji}D_{ji} - \widehat{\theta}_{\widehat{g}})$, $i = 1, \ldots, n$, and z_p is the pth quantile of standard normal distribution $\Phi(\cdot)$. Since the estimator $\widehat{\sigma}_g^2$ tends to underestimate σ_g^2, an alternative strategy is to utilize a bootstrap sample $\{W_i^*\}_{i=1}^n$ of $W_i = (D_i, Y_i)$'s. Following the derivation of [6], an approximate bootstrap variance of $\widehat{\theta}_{\widehat{g}}^*$ can be obtained as

$$V^*(\widehat{\theta}_{\widehat{g}}^*) = \frac{\left\{\sum_{\ell=0}^{1}(n_\ell - 1)(\widehat{\alpha}_\ell - \widehat{\theta}_{\widehat{g}}^2)\right\} + \widehat{\theta}_{\widehat{g}}(1-\widehat{\theta}_{\widehat{g}})}{n_0n_1} \tag{3.4}$$

with

$$\widehat{\alpha}_\ell = n_\ell^{-1}\sum_{i=1}^n\left\{n_{1-\ell}^{-1}\sum_{j=1}^n\widehat{\eta}_{ji}D_{1-\ell}^\ell(1-D_j)^\ell\right\}^2 D_i^\ell(1-D_i)^{1-\ell}, \quad \ell = 0, 1.$$

In the construction of confidence intervals, another preferred procedure is to use the re-sampling technique without relying on the asymptotic normality. The bootstrap variance and the frequency distribution of $\widehat{\theta}_{\widehat{g}}^* - \widehat{\theta}_{\widehat{g}}$ enable us to have a bootstrap-normal-type confidence interval and a bootstrap-quantile-type confidence interval:

$$\widehat{\theta}_{\widehat{g}} \pm z_{1-\alpha/2}\,\mathrm{se}^*(\widehat{\theta}_{\widehat{g}}^*) \quad \text{and} \quad (\widehat{\theta}_{\widehat{g}} - \gamma_{1,1-\alpha/2}^*(\widehat{\theta}_{\widehat{g}}^*), \widehat{\theta}_{\widehat{g}} - \gamma_{1,\alpha/2}^*(\widehat{\theta}_{\widehat{g}}^*)), \tag{3.5}$$

where $\mathrm{se}^*(\widehat{\theta}_{\widehat{g}}^*) = V^{*1/2}(\widehat{\theta}_{\widehat{g}}^*)$ and $\gamma_{1,p}^*(\widehat{\theta}_{\widehat{g}}^*)$ is the pth quantile of $\{(\widehat{\theta}_{\widehat{g}b}^* - \widehat{\theta}_{\widehat{g}})\}_{b=1}^B$ with $\widehat{\theta}_{\widehat{g}b}^*$ being computed based on the bth bootstrap sample $\{W_{bi}^*\}_{i=1}^n$ and B being the number of bootstrap replications. Let $P^*(\cdot) = P(\cdot|W_1, \ldots, W_n)$. The validity of confidence intervals in (3.5) is established in the next theorem.

Theorem 4. *Suppose that the assumptions in Theorem 1 or Theorem 2 are satisfied. Then,*

$$P\left(\frac{\widehat{\theta}_{\widehat{g}} - \theta_g}{\mathrm{se}^*(\widehat{\theta}_{\widehat{g}}^*)} \le z\right) \to \Phi(z) \text{ and} \tag{3.6}$$

$$P^*(\sqrt{n}(\widehat{\theta}_{\widehat{g}}^* - \widehat{\theta}_{\widehat{g}}) \le z) - P(\sqrt{n}(\widehat{\theta}_{\widehat{g}} - \theta_g) \le z) \xrightarrow{p} 0 \quad \text{for all } z \text{ as } n \to \infty. \tag{3.7}$$

Proof. See the Appendix. □

In practice, an estimator $\widehat{\theta}_{\widehat{g}^*}^*$ should be computed as a bootstrap analogue of $\widehat{\theta}_{\widehat{g}}$ rather than its asymptotically equivalent counterpart $\widehat{\theta}_{\widehat{g}}$. Another bootstrap-normal-type and bootstrap-quantile-type confidence interval for θ_g can be constructed by

$$\widehat{\theta}_{\widehat{g}} \pm z_{1-\alpha/2}\,\mathrm{se}^*(\widehat{\theta}_{\widehat{g}^*}^*) \quad \text{and} \quad (\widehat{\theta}_{\widehat{g}} - \gamma_{2,1-\alpha/2}^*(\widehat{\theta}_{\widehat{g}^*}^*), \widehat{\theta}_{\widehat{g}} - \gamma_{2,\alpha/2}^*(\widehat{\theta}_{\widehat{g}^*}^*)), \text{ respectively,} \tag{3.8}$$

where $\mathrm{se}^*(\widehat{\theta}_{\widehat{g}^*}^*)$ and $\gamma_{2,p}^*(\widehat{\theta}_{\widehat{g}^*}^*)$ are the standard deviation and the pth quantile of $\{(\widehat{\theta}_{\widehat{g}^*b}^* - \widehat{\theta}_{\widehat{g}})\}_{b=1}^B$.

3.3. Testing procedures

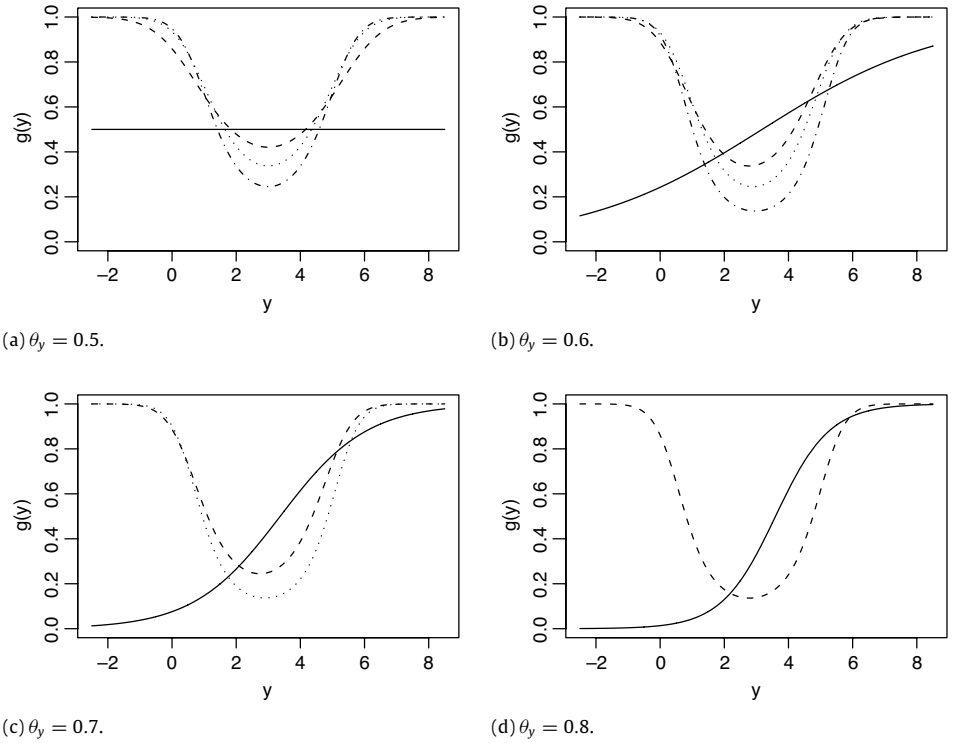
In the rest of this section, $\mathcal{R}_{g_F}(u)$ and $\mathcal{R}_{g_R}(u)$ are referred to as the ROC curves of transformations $g_F(Y_s)$ and $g_R(Y_s)$ in which $\mathcal{R}_{g_F}(u) \ge \mathcal{R}_{g_R}(u)$ uniformly over [0, 1]. The test statistic $\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{g_R}$ is naturally proposed to test the hypotheses:

$$\begin{cases} H_0 : \mathcal{R}_{g_F}(u) = \mathcal{R}_{g_R}(u) & \text{for all } u. \\ H_A : \mathcal{R}_{g_F}(u) > \mathcal{R}_{g_R}(u) & \text{for some } u. \end{cases} \tag{3.9}$$

In the above hypotheses, $(g_F(Y_s), g_R(Y_s)) = (g(Y), Y)$ is used to assess the impact of a transformation on the univariate marker Y. As for multiple markers, the setting $(g(Y), g_m^q(Y_s))$ for $m < q$ is mainly to screen out noninformative markers and $(g_m^q(Y_s), g_{msi}^q(Y_{\beta_s}))$ is considered to detect the possibility of dimension reduction in a higher-dimensional regression model. Moreover, the equality of the ROC curves of $g_F(Y_s) = g_{msi}^q(Y_{\beta_s})$ and $g_R(Y_s) = Y_{\beta_s}$ implies the monotonicity of $g_{msi}^q(y_{\beta_s})$ in y_{β_s} and, hence, the optimality of Y_{β_s}.

When the null hypothesis holds, one can easily show that $\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R} = \widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{g_F}$ and the asymptotic representation of $n^{1/2}(\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R})$ is $n^{-1/2}\sum_{i=1}^n(\Gamma_{\widehat{g}_Fi} - \Gamma_{\widehat{g}_Ri}) = o_p(1)$. Practitioners might be most likely to reject H_0 whenever

$$\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R} > z_{1-\alpha}\,\mathrm{se}(\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R}), \quad \text{where } \mathrm{se}(\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R}) = n^{-1}\left\{\sum_{i=1}^n(\widehat{\Gamma}_{\widehat{g}_Fi} - \widehat{\Gamma}_{\widehat{g}_Ri})^2\right\}^{1/2}. \tag{3.10}$$



(a) $\theta_y = 0.5$. (b) $\theta_y = 0.6$.

(c) $\theta_y = 0.7$. (d) $\theta_y = 0.8$.

Fig. 1. The curves of $g(y)$ with $\theta_Y = 0.5, 0.6, 0.7$, and 0.8, and $\Delta = 0$ (solid line), 0.1 (dashed line), 0.2 (dotted line), and 0.3 (dotted–dashed line).

The corresponding p-value is calculated as $p = \Phi((\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R})/\mathrm{se}(\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R}))$. Since the normal approximation for the conventional AUC test statistic of [5] is infeasible or problematic, a more accurate critical value could be obtained by taking into account higher-order approximation terms. Another way to overcome this difficulty is to employ the frequency distribution of bootstrap replications. The bootstrap test is then to reject H_0 if

$$\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R} > \gamma^*_{2,1-\alpha}(\widehat{\theta}^*_{\widehat{g}^*_F b} - \widehat{\theta}^*_{\widehat{g}^*_R b}) \tag{3.11}$$

and the bootstrap p-value can be calculated as

$$p^* = \frac{\sum_{b=1}^{B} I(\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R} > (\widehat{\theta}^*_{\widehat{g}^*_F b} - \widehat{\theta}^*_{\widehat{g}^*_R b}) - (\widehat{\theta}_{\widehat{g}_F} - \widehat{\theta}_{\widehat{g}_R}))}{B}.$$

4. Monte Carlo simulations

We conducted a class of simulations to assess the performances of the proposed estimation and inference procedures with a variety of sample combinations of n_0 and n_1. The simulation results were based on 2000 replications and the bootstrap inferences were drawn from 1000 bootstrap samples. This implementation enables us to obtain stable numerical results with differences in the second decimal digit of bias, empirical coverage probability, size, and power, and differences in the third decimal digit of standard deviation and standard error.

4.1. Simulation I — univariate marker

The univariate marker Y of a group $\{D = d\}$ was generated from a normal distribution with mean μ_d and variance σ_d^2, $d = 0, 1$. Different parameter values of $(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2)$ were chosen to produce the expected (θ_Y, θ_g) of $(Y, g(Y))$ and $\Delta = \theta_g - \theta_Y \in \{0, 0.1, 0.2, 0.3\}$. Fig. 1(a)–(d) display $g(y)$ and the corresponding values of $(\theta_Y, \theta_g, \Delta)$. The kernel estimator of θ_g was computed by (2.3) with the uniform kernel density. In our simulation experiments, the two-separate-bandwidth estimator was adopted because it outperforms the single-bandwidth and two-bandwidth ones, which were obtained from the cross-validation criterion.

Tables 1–3 exhibit the biases and the standard deviations of 2000 estimates, the standard errors based on the asymptotic the asymptotic formula below (3.3), the bootstrap standard errors based on the bootstrap variance in (3.4), and the empirical coverage probabilities of 0.95 confidence intervals with the sample sizes of 50, 100, or 200 for cases and controls. When

Table 1
The biases (b) and the standard deviations (sd) of 2000 estimates, the averages of standard errors (se, se^*), and the coverage probabilities (cp_n, cp_{bn}, cp_{bq}) based on 0.95 normal, bootstrap-normal, and bootstrap-quantile confidence intervals with $n_0 = n_1 = 50$.

θ_Y	θ_g	$b(\widehat{\theta}_{g_1})$	$sd(\widehat{\theta}_g)$	$se(\widehat{\theta}_g)$	$se^*(\widehat{\theta}_g)$	cp_n	cp_{bn}	cp_{bq}
0.5	0.5	0.000	0.0000	0.0000	0.0000	1.000	1.000	1.000
0.5	0.6	−0.001	0.0571	0.0564	0.0566	0.951	0.951	0.946
0.5	0.7	−0.002	0.0523	0.0522	0.0525	0.940	0.942	0.937
0.5	0.8	−0.001	0.0450	0.0448	0.0450	0.939	0.938	0.928
0.6	0.6	−0.001	0.0568	0.0563	0.0565	0.946	0.944	0.939
0.6	0.7	0.000	0.0524	0.0520	0.0523	0.942	0.943	0.937
0.6	0.8	0.001	0.0451	0.0445	0.0447	0.932	0.930	0.921
0.6	0.9	0.000	0.0324	0.0324	0.0326	0.924	0.927	0.896
0.7	0.7	−0.001	0.0511	0.0517	0.0520	0.942	0.941	0.936
0.7	0.8	0.000	0.0445	0.0446	0.0449	0.936	0.937	0.925
0.7	0.9	0.000	0.0332	0.0325	0.0327	0.920	0.920	0.893
0.8	0.8	0.001	0.0430	0.0432	0.0435	0.938	0.941	0.927
0.8	0.9	−0.002	0.0334	0.0327	0.0329	0.916	0.917	0.901
θ_Y	θ_g	$b(\widehat{\theta}_{\widehat{g}})$	$sd(\widehat{\theta}_{\widehat{g}})$	$se(\widehat{\theta}_{\widehat{g}})$	$se^*(\widehat{\theta}_{\widehat{g}})$	cp_n	cp_{bn}	cp_{bq}
0.5	0.5	0.082	0.0573	0.0566	0.0686	0.672	0.780	0.759
0.5	0.6	0.023	0.0572	0.0554	0.0651	0.914	0.946	0.913
0.5	0.7	0.009	0.0531	0.0514	0.0572	0.930	0.946	0.921
0.5	0.8	0.003	0.0449	0.0442	0.0475	0.932	0.951	0.926
0.6	0.6	0.015	0.0593	0.0556	0.0665	0.921	0.954	0.917
0.6	0.7	0.009	0.0535	0.0513	0.0570	0.926	0.950	0.920
0.6	0.8	0.005	0.0452	0.0441	0.0474	0.925	0.942	0.918
0.6	0.9	0.002	0.0329	0.0321	0.0337	0.913	0.929	0.880
0.7	0.7	0.000	0.0554	0.0515	0.0594	0.920	0.948	0.929
0.7	0.8	0.004	0.0446	0.0441	0.0475	0.928	0.940	0.908
0.7	0.9	0.001	0.0331	0.0321	0.0338	0.912	0.925	0.889
0.8	0.8	0.000	0.0445	0.0433	0.0479	0.934	0.948	0.936
0.8	0.9	0.000	0.0332	0.0324	0.0340	0.915	0.932	0.888

Table 2
The biases (b) and the standard deviations (sd) of 2000 estimates, the averages of standard errors (se, se^*), and the coverage probabilities (cp_n, cp_{bn}, cp_{bq}) based on 0.95 normal, bootstrap-normal, and bootstrap-quantile confidence intervals with $n_0 = n_1 = 100$.

θ_Y	θ_g	$b(\widehat{\theta}_g)$	$sd(\widehat{\theta}_g)$	$se(\widehat{\theta}_g)$	$se^*(\widehat{\theta}_g)$	cp_n	cp_{bn}	cp_{bq}
0.5	0.5	0.000	0.0000	0.0000	0.0000	1.000	1.000	1.000
0.5	0.6	−0.001	0.0398	0.0399	0.0400	0.944	0.945	0.941
0.5	0.7	−0.001	0.0372	0.0369	0.0370	0.943	0.945	0.941
0.5	0.8	0.000	0.0316	0.0316	0.0317	0.946	0.947	0.939
0.6	0.6	0.000	0.0398	0.0398	0.0399	0.953	0.953	0.948
0.6	0.7	0.000	0.0362	0.0368	0.0369	0.951	0.952	0.946
0.6	0.8	−0.001	0.0313	0.0317	0.0318	0.950	0.948	0.942
0.6	0.9	0.000	0.0226	0.0231	0.0232	0.942	0.942	0.932
0.7	0.7	0.000	0.0369	0.0365	0.0366	0.947	0.947	0.941
0.7	0.8	0.000	0.0323	0.0316	0.0317	0.943	0.945	0.935
0.7	0.9	0.000	0.0229	0.0232	0.0233	0.940	0.943	0.933
0.8	0.8	0.000	0.0312	0.0307	0.0308	0.948	0.947	0.936
0.8	0.9	0.000	0.0227	0.0232	0.0233	0.937	0.939	0.928
θ_Y	θ_{g1}	$b(\widehat{\theta}_{\widehat{g}})$	$sd(\widehat{\theta}_{\widehat{g}})$	$se(\widehat{\theta}_{\widehat{g}})$	$se^*(\widehat{\theta}_{\widehat{g}})$	cp_n	cp_{bn}	cp_{bq}
0.5	0.5	0.061	0.0404	0.0404	0.0492	0.656	0.764	0.769
0.5	0.6	0.011	0.0407	0.0396	0.0460	0.929	0.955	0.936
0.5	0.7	0.003	0.0379	0.0367	0.0400	0.939	0.956	0.935
0.5	0.8	0.002	0.0315	0.0315	0.0331	0.945	0.960	0.937
0.6	0.6	0.008	0.0410	0.0396	0.0465	0.936	0.966	0.940
0.6	0.7	0.005	0.0365	0.0366	0.0398	0.938	0.960	0.939
0.6	0.8	0.001	0.0315	0.0315	0.0332	0.941	0.951	0.939
0.6	0.9	0.001	0.0226	0.0230	0.0237	0.934	0.944	0.922
0.7	0.7	−0.001	0.0383	0.0365	0.0405	0.937	0.954	0.947
0.7	0.8	0.002	0.0321	0.0315	0.0331	0.938	0.954	0.931
0.7	0.9	0.001	0.0230	0.0231	0.0238	0.938	0.948	0.920
0.8	0.8	−0.001	0.0319	0.0308	0.0326	0.944	0.958	0.943
0.8	0.9	0.001	0.0227	0.0231	0.0239	0.936	0.947	0.925

the sample size is moderate, the performances of $\widehat{\theta}_g$, which cannot be obtained from real data, and $\widehat{\theta}_{\widehat{g}}$ are very similar. As expected, the biases and the standard deviations tend to be smaller for a larger θ_g or sample size. With small sample size combinations (e.g. $(n_0, n_1) = (50, 50)$), one can see that the averages of 2000 standard errors are somewhat smaller than

Table 3
The biases (b) and the standard deviations (sd) of 2000 estimates, the averages of standard errors (se, se^*), and the coverage probabilities (cp_n, cp_{bn}, cp_{bq}) based on 0.95 normal, bootstrap-normal, and bootstrap-quantile confidence intervals with $n_0 = n_1 = 200$.

θ_Y	θ_g	$b(\widehat{\theta_g})$	$sd(\widehat{\theta_g})$	$se(\widehat{\theta_g})$	$se^*(\widehat{\theta_g})$	cp_n	cp_{bn}	cp_{bq}
0.5	0.5	0.000	0.0000	0.0000	0.0000	1.000	1.000	1.000
0.5	0.6	0.001	0.0280	0.0282	0.0282	0.949	0.950	0.946
0.5	0.7	0.000	0.0262	0.0261	0.0261	0.947	0.948	0.943
0.5	0.8	0.000	0.0221	0.0224	0.0224	0.948	0.949	0.945
0.6	0.6	0.000	0.0284	0.0281	0.0282	0.945	0.945	0.945
0.6	0.7	0.001	0.0252	0.0261	0.0261	0.955	0.956	0.952
0.6	0.8	0.000	0.0227	0.0224	0.0224	0.947	0.947	0.944
0.6	0.9	0.000	0.0162	0.0164	0.0165	0.946	0.945	0.940
0.7	0.7	0.001	0.0259	0.0258	0.0258	0.946	0.947	0.942
0.7	0.8	−0.001	0.0224	0.0224	0.0225	0.952	0.952	0.947
0.7	0.9	0.000	0.0163	0.0165	0.0165	0.950	0.949	0.944
0.8	0.8	0.000	0.0222	0.0217	0.0217	0.947	0.945	0.942
0.8	0.9	0.000	0.0163	0.0164	0.0164	0.943	0.944	0.936
θ_Y	θ_g	$b(\widehat{\theta_{\widehat{g}}})$	$sd(\widehat{\theta_{\widehat{g}}})$	$se(\widehat{\theta_{\widehat{g}}})$	$se^*(\widehat{\theta_{\widehat{g}}})$	cp_n	cp_{bn}	cp_{bq}
0.5	0.5	0.046	0.0292	0.0287	0.0351	0.618	0.740	0.768
0.5	0.6	0.007	0.0285	0.0281	0.0318	0.935	0.957	0.940
0.5	0.7	0.003	0.0264	0.0260	0.0277	0.943	0.959	0.944
0.5	0.8	0.001	0.0220	0.0223	0.0231	0.944	0.950	0.941
0.6	0.6	0.003	0.0292	0.0281	0.0323	0.934	0.963	0.944
0.6	0.7	0.004	0.0254	0.0260	0.0276	0.950	0.963	0.954
0.6	0.8	0.001	0.0225	0.0223	0.0231	0.946	0.951	0.944
0.6	0.9	0.001	0.0162	0.0164	0.0167	0.943	0.950	0.935
0.7	0.7	0.000	0.0265	0.0258	0.0275	0.941	0.955	0.951
0.7	0.8	0.001	0.0228	0.0223	0.0232	0.945	0.952	0.945
0.7	0.9	0.000	0.0163	0.0164	0.0168	0.945	0.950	0.941
0.8	0.8	−0.001	0.0223	0.0217	0.0224	0.947	0.952	0.950
0.8	0.9	0.001	0.0162	0.0164	0.0167	0.943	0.950	0.936

Table 4.1
The biases (b) and the standard deviations (sd) of 2000 estimates, the averages of standard errors (se, se^*), and the coverage probabilities (cp_n, cp_{bn}, cp_{bq}) based on 0.95 normal, bootstrap-normal, and bootstrap-quantile confidence intervals with three sample sizes.

$n_0 = n_1$	θ_y	θ_g	$b(\widehat{\theta_Y})$	$sd(\widehat{\theta_Y})$	$se(\widehat{\theta_Y})$	$se^*(\widehat{\theta_Y})$	cp_n	cp_{bn}	cp_{bq}
50	0.5	0.5	0.000	0.0576	0.0577	0.0580	0.947	0.949	0.938
	0.5	0.6	0.000	0.0589	0.0582	0.0585	0.942	0.941	0.937
	0.5	0.7	−0.001	0.0606	0.0597	0.0600	0.941	0.942	0.935
	0.5	0.8	−0.002	0.0610	0.0622	0.0623	0.949	0.949	0.940
100	0.5	0.5	0.001	0.0408	0.0408	0.0409	0.943	0.943	0.941
	0.5	0.6	0.000	0.0412	0.0412	0.0412	0.952	0.954	0.948
	0.5	0.7	0.001	0.0417	0.0423	0.0423	0.954	0.953	0.949
	0.5	0.8	−0.001	0.0452	0.0441	0.0441	0.942	0.943	0.937
200	0.5	0.5	−0.001	0.0285	0.0289	0.0289	0.954	0.955	0.951
	0.5	0.6	0.001	0.0296	0.0291	0.0292	0.949	0.947	0.950
	0.5	0.7	0.001	0.0300	0.0299	0.0299	0.945	0.946	0.939
	0.5	0.8	0.000	0.0313	0.0312	0.0312	0.949	0.946	0.941

the standard deviations of their estimates whereas the bootstrap standard errors are relatively greater. It is further detected from these tables that the estimates of standard deviations tend to be close to the true values as the sample size increases. Except very poor coverage rates for $\theta_g = 0.5$ and noticeably lower ones for $\theta_g = 0.9$ with $(n_0, n_1) = (50, 50)$, which has been confirmed by [18], the majority of the empirical coverage probabilities are found near the nominal level of 0.95. In addition, the bootstrap-normal-type confidence interval in (3.8) is detected to have better coverage probabilities than the others. One can show that $\sqrt{n}(\widehat{\theta_{\widehat{g}}} - 0.5)$ does not converge in probability to zero for the case of $\theta_Y = \theta_g = 0.5$. Relative to the standard deviation, the bias is not ignorable and is quite likely to be the main cause for the poor performance. To avoid this drawback, we would suggest first examining whether the constructed confidence intervals of θ_Y (Table 4.1) contain the value of 0.5 and then performing the independence test for Y and D (Table 4.2).

In Table 5, the estimated sizes and powers are provided for the hypotheses of the identical ROC curve of $g(Y)$ and Y versus the uniform alternatives in (3.9). It is found that the estimated sizes of our nonparametric area test are generally within the significance level of 0.05 except $\theta_Y = \theta_g = 0.5$. Since the size is often overestimated in this circumstance, the Kolmogorov–Smirnov goodness of fit statistic

$$\sup_y \left| \sum_{d=0}^{1} (-1)^{d+1} n_d^{-1} \sum_{i=1}^{n} I(Y_i \le y) D_i^d (1 - D_i)^{1-d} \right|$$

Table 4.2
Estimated sizes (α) and powers (β) of Kolmogorov–Smirnov goodness of fit test with the significant level of 0.05 and three sample sizes.

$n_0 = n_1$		50	100	200
θ_Y	θ_g	α	α	α
0.5	0.5	0.037	0.035	0.047
θ_Y	θ_g	β	β	β
0.5	0.6	0.082	0.123	0.391
0.5	0.7	0.351	0.739	0.997
0.5	0.8	0.910	0.999	1.000

Table 5
Estimated sizes (α, α^*) and powers (β, β^*) based on the normal and bootstrap testing procedures with the significant level of 0.05 and three sample sizes.

$n_0 = n_1$		50		100		200	
θ_Y	θ_g	α	α^*	α	α^*	α	α^*
0.5	0.5	0.177	0.148	0.224	0.152	0.259	0.168
0.6	0.6	0.044	0.015	0.053	0.014	0.044	0.004
0.7	0.7	0.022	0.006	0.023	0.000	0.030	0.000
0.8	0.8	0.016	0.003	0.023	0.000	0.023	0.000
θ_Y	θ_g	β	β^*	β	β^*	β	β^*
0.5	0.6	0.411	0.348	0.616	0.495	0.872	0.762
0.5	0.7	0.853	0.786	0.986	0.965	1.000	1.000
0.5	0.8	0.995	0.987	1.000	1.000	1.000	1.000
0.6	0.7	0.489	0.412	0.811	0.716	0.973	0.939
0.6	0.8	0.926	0.903	0.998	0.994	1.000	1.000
0.6	0.9	0.999	0.998	1.000	1.000	1.000	1.000
0.7	0.8	0.594	0.522	0.888	0.834	0.994	0.988
0.7	0.9	0.971	0.961	1.000	1.000	1.000	1.000
0.8	0.9	0.680	0.643	0.947	0.929	1.000	1.000

could be used to test the hypotheses (Table 4.2). Across several sets of (θ_Y, θ_g) and (n_0, n_1), the variation in the power of the area test is detected to be an increasing function of Δ, θ_Y and θ_g for each fixed Δ, and $n = n_0 + n_1$. The same as the conclusion drawn in the former literature, the applied test usually has a higher power under the uniform alternatives.

4.2. Simulation $\mathbb{I}$ — multiple markers

The finite sample properties of our estimation for the accuracy measure of an optimal composite marker and the related inference procedures of interest were investigated through three regression models. The estimator of $g_m^q(y)$ with the uniform kernel density and that of $g_{msi}^q(y_\beta)$ with the Epanechnikov kernel density were computed by (2.1) and (2.4), respectively. The corresponding AUC estimates in (2.3) and (2.5) were obtained in the same way with their bandwidths being chosen from the cross-validation criteria in (2.6) and (2.7).

Let $Y = (Z_1 + Z_4, Z_2 + Z_5, Z_3)^\top$ and $Y_s = (Y_1, Y_2)^\top$ with (Z_1, Z_2, Z_3) being a multivariate normal distribution with mean (0.5, 0.2,2), standard deviation (1, 1, 2), and correlation coefficient 0.7. In addition, Z_4 and Z_5 were independently generated from gamma distributions with the parameters (α, β) of (5, 1) or (3, 5). The first considered model

$$\text{M1. } P(D = 1 \mid y) = \frac{\exp(4 - y_1 + 2y_2)}{1 + \exp(4 - y_1 + 2y_2)} \quad \text{with } (P(D = 1), \theta_{g_q}) = (0.530, 0.906)$$

was mainly designed to investigate the performance of our test rules for the hypotheses in (3.9) with $(g(Y), g_2^3(Y_s))$, $(g_2^3(Y_s), g_{2si}^3(Y_{\beta_s}))$, and $(g_{2si}^3(Y_{\beta_s}), Y_{\beta_s})$ being considered for $(g_F(Y_s), g_R(Y_s))$. It is implied from model M1 that $\theta_g = \theta_{g_2^3} = \theta_{g_{2si}^3} = \theta_{Y_{\beta_s}}$. Table 6.1 reveals that the biases in $\widehat{\theta}_{\widehat{g}_2^3}$, $\widehat{\theta}_{\widehat{g}_{2si}^3}$, and $\widehat{\theta}_{Y_{\widehat{\beta}_s}}$ are not apparent and their standard deviations are close to each other under moderate sample sizes. In contrast, the estimator $\widehat{\theta}_{\widehat{g}}$ is much easier to underestimate θ_g and its variation is relatively large compared to that of $\widehat{\theta}_{\widehat{g}_2^3}$. One can see that the empirical coverage probabilities of normal-type confidence intervals are somewhat lower than 0.95 but those of bootstrap-normal-type and bootstrap-quantile-type ones are relatively close to the assigned nominal level. Under model M1, the estimated sizes are all found to be within the significance level of 0.05 (Table 6.2).

The second scenario was provided to study the influence of a misspecified linear predictor on the AUC estimator and the related inference procedures. For this purpose, we specify $Y = (Y_{10}, Y_{20}, Y_{10}Y_{20})^\top$ and $Y_s = (Y_1, Y_2)^\top$ with (Y_{10}, Y_{20}) following a bivariate normal distribution with mean (0.5, 0.2), standard deviation (1, 1), and correlation coefficient 0.3. The logistic regression model is designed as

$$\text{M2. } P(D = 1 \mid y_1, y_2) = \frac{\exp(1 - y_1 + 1.5y_2 - y_1y_2)}{1 + \exp(1 - y_1 + 1.5y_2 - y_1y_2)} \quad \text{with } (P(D = 1), \theta_g) = (0.561, 0.803).$$

Table 6.1
The biases (b) and the standard deviations (sd) of 2000 estimates, the averages of standard errors (se, se^*), and the coverage probabilities (cp_n, cp_{bn}, cp_{bq}) based on 0.95 normal-type, bootstrap-normal-type, and bootstrap-quantile-type confidence intervals with three sample sizes (n) under model M1.

n	$b(\widehat{\theta}_{g_2^3})$	$sd(\widehat{\theta}_{g_2^3})$	$se(\widehat{\theta}_{g_2^3})$	$se^*(\widehat{\theta}_{g_2^3})$	cp_n	cp_{bn}	cp_{bq}
300	0.000	0.0168	0.0166	0.0167	0.940	0.944	0.932
400	0.000	0.0141	0.0144	0.0144	0.946	0.946	0.938
600	0.000	0.0119	0.0117	0.0118	0.945	0.944	0.936
n	$b(\widehat{\theta}_{\widetilde{g}})$	$sd(\widehat{\theta}_{\widetilde{g}})$	$se(\widehat{\theta}_{\widetilde{g}})$	$se^*(\widehat{\theta}_{\widetilde{g}})$	cp_n	cp_{bn}	cp_{bq}
300	−0.009	0.0187	0.0176	0.0203	0.922	0.962	0.959
400	−0.008	0.0157	0.0152	0.0172	0.930	0.961	0.971
600	0.006	0.0122	0.0123	0.0136	0.942	0.965	0.966
n	$b(\widehat{\theta}_{\widetilde{g}_2^3})$	$sd(\widehat{\theta}_{\widetilde{g}_2^3})$	$se(\widehat{\theta}_{\widetilde{g}_2^3})$	$se^*(\widehat{\theta}_{\widetilde{g}_2^3})$	cp_n	cp_{bn}	cp_{bq}
300	−0.002	0.0179	0.0169	0.0188	0.932	0.955	0.932
400	−0.002	0.0150	0.0147	0.0160	0.945	0.967	0.949
600	−0.002	0.0124	0.0120	0.0128	0.942	0.956	0.947
n	$b(\widehat{\theta}_{\widetilde{g}_{2si}^3})$	$sd(\widehat{\theta}_{\widetilde{g}_{2si}^3})$	$se(\widehat{\theta}_{\widetilde{g}_{2si}^3})$	$se^*(\widehat{\theta}_{\widetilde{g}_{2si}^3})$	cp_n	cp_{bn}	cp_{bq}
300	0.001	0.0169	0.0164	0.0165	0.928	0.943	0.917
400	0.001	0.0141	0.0142	0.0143	0.934	0.946	0.929
600	0.001	0.0120	0.0117	0.0117	0.939	0.944	0.930
n	$b(\widehat{\theta}_{Y_{\widehat{\beta}_s}})$	$sd(\widehat{\theta}_{Y_{\widehat{\beta}_s}})$	$se(\widehat{\theta}_{Y_{\widehat{\beta}_s}})$	$se^*(\widehat{\theta}_{Y_{\widehat{\beta}_s}})$	cp_n	cp_{bn}	cp_{bq}
300	0.001	0.0168	0.0165	0.0166	0.938	0.939	0.931
400	0.001	0.0140	0.0143	0.0143	0.940	0.940	0.935
600	0.000	0.0119	0.0117	0.0118	0.943	0.943	0.940

Table 6.2
Estimated sizes (α, α^*) based on the normal and bootstrap testing procedures with the significant level of 0.05 and three sample sizes (n) under model M1.

n	300		400		600	
H_0	α	α^*	α	α^*	α	α^*
$\mathcal{R}_g(u) = \mathcal{R}_{g_2^3}(u)$	0.016	0.002	0.017	0.004	0.019	0.004
$\mathcal{R}_{g_2^3}(u) = \mathcal{R}_{g_{2si}^3}(u)$	0.009	0.022	0.006	0.023	0.005	0.010
$\mathcal{R}_{g_{2si}^3}(u) = \mathcal{R}_{Y_{\beta_s}}(u)$	0.052	0.035	0.057	0.023	0.058	0.035

The scoring functions $g_{3si}^3(Y_\beta)$ and $g_{2si}^3(Y_{\beta_s})$ and their accuracy measures were computed in simulated data. As expected, the performance of $\widehat{\theta}_{\widetilde{g}_{2si}^3}$ and its inference procedure are particularly poor, whereas the other ones perform rather well in moderate sample sizes (Table 7.1). Under the equality of two ROC curves: $\mathcal{R}_g(u) = \mathcal{R}_{g_{3si}^3}(u)$, the estimated sizes in Table 7.2 are fairly close to the nominal size at 0.05 significance level but those of the bootstrap test are slightly overestimated. As for the hypothesis of $\mathcal{R}_g(u) = \mathcal{R}_{g_{2si}^3}(u)$, a higher power is usually associated with a larger sample size.

In the last numerical experiment, $Y = (Y_1, Y_2)^\top$ was specified with $(Y_1, Y_2) = (0.3W_1 + W_2, 0.3W_1 + W_3)$ and W_1, W_2, and W_3 being independently generated from beta distributions with the corresponding parameters (5, 5) or (8, 3) or (2, 7). The model

$$\text{M3. } P(D = 1 \mid y_1, y_2) = \frac{2.35(0.8y_1 + 0.8y_2)^4}{\exp\{(0.8y_1 + 0.8y_2)^5\}} \quad \text{with } (P(D = 1), \theta_{g_q}) = (0.684, 0.762)$$

was employed to explore the impact of a nonmonotonic link function on inferences. To provide an illustration of model misspecification, a logistic regression with linear predictor $Y_{\beta_{logit}}$ was considered as one possible working model. Since the logit link function is misspecified, the poor accuracy of $\widehat{\theta}_{Y_{\widehat{\beta}_{logit}}}$ and the poor performance of its inference procedure are detected (Table 8.1). It is further indicated from Table 8.2 that the sizes are fairly close to the nominal size of 0.05 under $\mathcal{R}_g(u) = \mathcal{R}_{g_{2si}^2}(u)$ and the powers are very high for testing $\mathcal{R}_{g_{2si}^2}(u) = \mathcal{R}_{Y_{\beta_{logit}}}(u)$.

5. Data examples

The first theme in our data analyses is to assess the necessity of transformation on the individual biomarkers in the early detection of binary disease status. The second objective is to improve the classification capability through seeking an optimal composite biomarker. Hypotheses for the equality of $\mathcal{R}_g(u)$ and $\mathcal{R}_{g_m^q}(u)$ are tested sequentially according to the order $m = (q-1), \ldots, 1$, with all possible subsets Y_s of size m. In the preliminary data analyses, some biomarkers might be identified to be noninformative and, hence, $g_m^q(y_s)$ can be directly specified to be the transformation of the other ones.

Table 7.1
The biases (b) and the standard deviations (sd) of 2000 estimates, the averages of standard errors (se, se^*), and the coverage probabilities (cp_n, cp_{bn}, cp_{bq}) based on 0.95 normal-type, bootstrap-normal-type, and bootstrap-quantile-type confidence intervals with three sample sizes (n) under model M2.

n	$b(\widehat{\theta}_g)$	$sd(\widehat{\theta}_g)$	$se(\widehat{\theta}_g)$	$se^*(\widehat{\theta}_g)$	cp_n	cp_{bn}	cp_{bq}
300	0.000	0.0250	0.0248	0.0249	0.941	0.944	0.941
400	0.001	0.0218	0.0214	0.0215	0.943	0.944	0.937
600	0.000	0.0175	0.0175	0.0176	0.949	0.951	0.946
n	$b(\widehat{\theta}_{\widehat{g}})$	$sd(\widehat{\theta}_{\widehat{g}})$	$se(\widehat{\theta}_{\widehat{g}})$	$se^*(\widehat{\theta}_{\widehat{g}})$	cp_n	cp_{bn}	cp_{bq}
300	0.003	0.0253	0.0247	0.0280	0.938	0.961	0.946
400	0.004	0.0222	0.0214	0.0236	0.925	0.947	0.932
600	0.001	0.0179	0.0175	0.0189	0.940	0.958	0.950
n	$b(\widehat{\theta}_{\widehat{g}^3_{3si}})$	$sd(\widehat{\theta}_{\widehat{g}^3_{3si}})$	$se(\widehat{\theta}_{\widehat{g}^3_{3si}})$	$se^*(\widehat{\theta}_{\widehat{g}^3_{3si}})$	cp_n	cp_{bn}	cp_{bq}
300	0.004	0.0250	0.0245	0.0246	0.930	0.931	0.923
400	0.003	0.0220	0.0213	0.0213	0.932	0.931	0.928
600	0.001	0.0177	0.0175	0.0175	0.940	0.941	0.935
n	$b(\widehat{\theta}_{\widehat{g}^3_{2si}})$	$sd(\widehat{\theta}_{\widehat{g}^3_{2si}})$	$se(\widehat{\theta}_{\widehat{g}^3_{2si}})$	$se^*(\widehat{\theta}_{\widehat{g}^3_{2si}})$	cp_n	cp_{bn}	cp_{bq}
300	−0.046	0.0290	0.0274	0.0275	0.623	0.606	0.642
400	−0.047	0.0246	0.0238	0.0239	0.512	0.490	0.529
600	−0.050	0.0197	0.0195	0.0195	0.270	0.258	0.284

Table 7.2
Estimated sizes (α, α^*) and powers (β, β^*) based on the normal and bootstrap testing procedures with the significant level of 0.05 and three sample sizes (n) under model M2.

n	300		400		600	
H_0	α	α^*	α	α^*	α	α^*
$\mathcal{R}_g(u) = \mathcal{R}_{g^3_{3si}}(u)$	0.009	0.048	0.011	0.059	0.007	0.055
	β	β^*	β	β^*	β	β^*
$\mathcal{R}_g(u) = \mathcal{R}_{g^3_{2si}}(u)$	0.759	0.806	0.890	0.926	0.978	0.986

Table 8.1
The biases (b) and the standard deviations (sd) of 2000 estimates, the averages of standard errors (se, se^*), and the coverage probabilities (cp_n, cp_{bn}, cp_{bq}) based on 0.95 normal-type, bootstrap-normal-type, and bootstrap-quantile-type confidence intervals with three sample sizes (n) under model M3.

n	$b(\widehat{\theta}_g)$	$sd(\widehat{\theta}_g)$	$se(\widehat{\theta}_g)$	$se^*(\widehat{\theta}_g)$	cp_n	cp_{bn}	cp_{bq}
300	−0.001	0.0312	0.0308	0.0309	0.946	0.948	0.943
400	−0.000	0.0268	0.0266	0.0267	0.946	0.946	0.940
600	−0.000	0.0221	0.0218	0.0218	0.948	0.949	0.948
n	$b(\widehat{\theta}_{\widehat{g}})$	$sd(\widehat{\theta}_{\widehat{g}})$	$se(\widehat{\theta}_{\widehat{g}})$	$se^*(\widehat{\theta}_{\widehat{g}})$	cp_n	cp_{bn}	cp_{bq}
300	0.007	0.0327	0.0302	0.0346	0.906	0.942	0.919
400	0.006	0.0288	0.0263	0.0297	0.913	0.943	0.926
600	0.004	0.0229	0.0216	0.0241	0.926	0.955	0.945
n	$b(\widehat{\theta}_{\widehat{g}^2_{2si}})$	$sd(\widehat{\theta}_{\widehat{g}^2_{2si}})$	$se(\widehat{\theta}_{\widehat{g}^2_{2si}})$	$se^*(\widehat{\theta}_{\widehat{g}^2_{2si}})$	cp_n	cp_{bn}	cp_{bq}
300	0.012	0.0327	0.0300	0.0301	0.893	0.949	0.883
400	0.010	0.0277	0.0261	0.0261	0.902	0.953	0.896
600	0.007	0.0223	0.0214	0.0215	0.924	0.955	0.916
n	$b(\widehat{\theta}_{Y_{\widehat{\beta}_s}})$	$sd(\widehat{\theta}_{Y_{\widehat{\beta}_s}})$	$se(\widehat{\theta}_{Y_{\widehat{\beta}_s}})$	$se^*(\widehat{\theta}_{Y_{\widehat{\beta}_s}})$	cp_n	cp_{bn}	cp_{bq}
300	−0.127	0.0399	0.0385	0.0383	0.093	0.102	0.092
400	−0.129	0.0339	0.0334	0.0336	0.029	0.031	0.029
600	−0.131	0.0278	0.0275	0.0280	0.002	0.003	0.003

Table 8.2
Estimated sizes (α, α^*) and powers (β, β^*) based on the normal and bootstrap testing procedures with the significant level of 0.05 and three sample sizes (n) under model M3.

n	300		400		600	
H_0	α	α^*	α	α^*	α	α^*
$\mathcal{R}_g(u) = \mathcal{R}_{g^2_{2si}}(u)$	0.014	0.038	0.011	0.029	0.012	0.039
	β	β^*	β	β^*	β	β^*
$\mathcal{R}_{g^2_{2si}}(u) = \mathcal{R}_{Y_{\beta_{logit}}}(u)$	0.987	0.985	0.997	0.998	1.000	1.000

Table 9
The estimates for AUCs of six biomarkers and their optimal transformations, the standard errors (se and se^*) and the p-values (p and p^*) based on the normal and bootstrap testing procedures.

Biomarker	$\widehat{\theta}_{Y_s}$	$se(\widehat{\theta}_{Y_s})$	$se^*(\widehat{\theta}_{Y_s})$	$\widehat{\theta}_{\widehat{g}_1^6}$	$se(\widehat{\theta}_{\widehat{g}_1^6})$	$se^*(\widehat{\theta}_{\widehat{g}_1^6})$	p	p^*
gtt	0.792	0.0169	0.0168	0.793	0.0169	0.0169	0.348	0.441
dbp	0.582	0.0219	0.0222	0.615	0.0212	0.0224	0.021	0.080
tsft	0.551	0.0228	0.0228	0.638	0.0202	0.0200	0.000	0.002
si	0.539	0.0222	0.0221	0.644	0.0189	0.0199	0.000	0.000
bmi	0.684	0.0193	0.0199	0.689	0.0192	0.0213	0.328	0.427
dpf	0.606	0.0216	0.0216	0.618	0.0216	0.0233	0.043	0.316

When there are more than one subsets of biomarkers with the same size are tested to be appropriate, the transformation of that possessing the greatest p-value will be suggested in application. Specifically, a related interest is to explore whether $g_m^q(y_s)$ can be simplified to the form $g_{msi}^q(y_{\beta_s})$ or even y_{β_s} based on the monotonicity of $g_{msi}^q(y_{\beta_s})$ in y_{β_s} or some widely used generalized linear models.

5.1. Application to a study of Pima–Indian diabetes

The analyzed data were collected from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDKD). A total of 798 Pima–Indian women, who were at least 21 years of age at the time of index examination, were chosen for forecasting the onset of diabetes between one and five years from the examination. Measurements taken included the number of times pregnant, age, and six biomarkers: plasma glucose concentration at two hours in an oral glucose tolerance (*gtt*), diastolic blood pressure (*dbp*), triceps skin fold thickness (*tsft*), 2-h serum insulin (*si*), body mass index (*bmi*), and diabetes pedigree function (*dpf*). Patients with zero *bmi* or *gtt* were excluded to ensure better quality of data. Out of 764 female patients, 264 were diagnosed with diabetes according to the World Health Organization criteria mentioned in [22] while the remaining 500 patients did not.

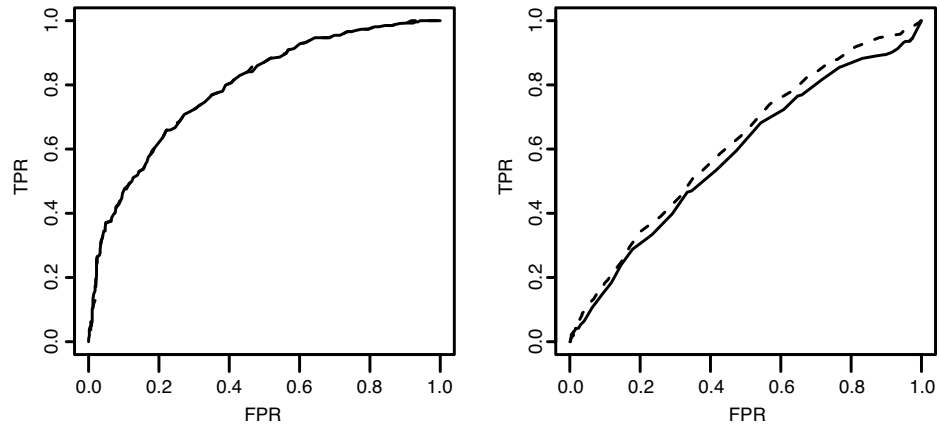
One can see from Table 9 that the biomarkers of *gtt* and *bmi* have acceptable and near acceptable discriminating accuracies whereas the other ones are relatively poor especially both of *tsft* and *si*. The optimal transformations of *tsft* and *si* have significantly higher AUCs than the original measurements and have better classification capabilities than *dbp* and *dpf*. However, no transformation is required on *gtt*, *bmi*, and *dpf*. These findings can be evidenced through the estimated ROC curves in Fig. 2(a)–(f). With the use of $g_F(Y_s) = g(gtt, dbp, tsft, si, bmi, dpf)$ and $g_R(Y_s) = g_4^6(gtt, dbp, bmi, dpf)$ for the hypothesis $\mathcal{R}_{g_F}(u) = \mathcal{R}_{g_R}(u)$, $g_4^6(gtt, dbp, bmi, dpf)$ is shown to be an adequate composite marker among all possible combinations of these six biomarkers ($p = 0.999$; $p^* = 0.993$). An excellent discriminating ability of $g_4^6(gtt, dbp, bmi, dpf)$ is further implied from the estimate (standard error; bootstrap standard error) 0.845 (0.0143; 0.0148) of its accuracy measure $\theta_{g_4^6}$. Compared with the estimated AUC of *gtt*, about a six percent increment is obtained in $g_4^6(gtt, dbp, bmi, dpf)$.

In this data analysis, a SIM with an unspecified link function and linear predictor $Y_{\beta_s} = gtt + \beta_{2s}dbp + \beta_{3s}bmi + \beta_{4s}dpf$ was employed to characterize the relationship between diabetes mellitus and baseline biomarkers (gtt, dbp, bmi, dpf). The estimated scoring function $\widehat{g}_{4si}^6(gtt - 0.122dbp + 2.203bmi + 34.448dpf)$ is obtained with $\widehat{\theta}_{\widehat{g}_{4si}^6}$ being 0.831 (0.0150; 0.0150). There is insufficient evidence to show its optimality from the p-value $p = 0.012$, while the bootstrap p-value $p^* = 0.073$ provides mild indications that it might be an optimal biomarker. In addition, both logistic and probit models were adopted as possible working models. Using the generalized estimating equation approach, their linear predictors are estimated as $Y_{\widehat{\beta}_{slogit}} = -8.103 + 0.039gtt - 0.008dbp + 0.085bmi + 0.869dpf$ and $Y_{\widehat{\beta}_{sprobit}} = -4.748 + 0.023gtt - 0.005dbp + 0.050bmi + 0.439dpf$, and the estimates of $\theta_{y_{\beta_{slogit}}}$ and $\theta_{y_{\beta_{sprobit}}}$ are computed to be 0.828 (0.0151; 0.0151) and 0.828 (0.0151; 0.0152). The p-values (p; p^*)(0.008; 0.034) for testing $\mathcal{R}_{g_{4si}^6}(u) = \mathcal{R}_{Y_{\beta_{slogit}}}(u)$ and (0.008; 0.036) for testing $\mathcal{R}_{g_{4si}^6}(u) = \mathcal{R}_{Y_{\beta_{sprobit}}}(u)$ indicate that the considered parametric models might be misspecified. We note that $\widehat{\theta}_{\widehat{g}_{4si}^6}$ is slightly higher than the sigmoid AUC estimate (0.825 with a standard error of 0.023), which was computed based on a linear combination of biomarkers and some demographic variables, of [13]. The same as a univariate marker, our analysis results show that $g_{4si}^6(y_{\beta_s})$ might be strictly increasing in y_{β_s} or a more complex nonlinear model.

5.2. Application to a study of liver disorder

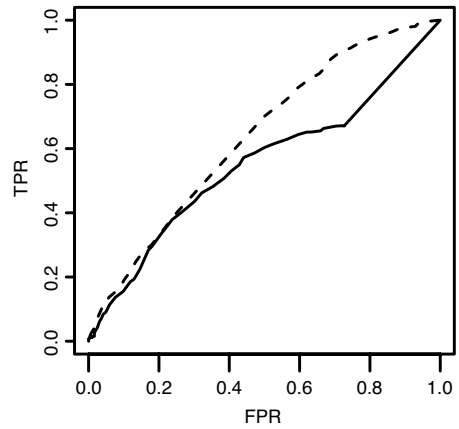
The liver disorder data were obtained from the BUPA medical research Ltd database donated by Richard S. Forsyth. It consists of 345 patients in which 145 samples have liver disorders. For all patients, the results of five blood tests are available on a continuous scale as follows: mean corpuscular volume (mcv), alkaline phosphotase (*alphos*), alamine aminotransferase (sgpt), aspartate aminotransferase (*sgot*), and γ-glutamyl transpeptidase (γgt), and the number of alcoholic beverages drunk per day (*drink*). The details of data information can also be found in [26].

From Table 10 and Fig. 3(a)–(e), it appears that the classification abilities of these five blood tests are poor. Furthermore, no apparent improvement is detected in the accuracy measures of the univariate transformations of *alphos*, *sgot*, and γgt. Although the AUCs of the optimal transformations of mcv and *sgpt* are significantly higher than those of the

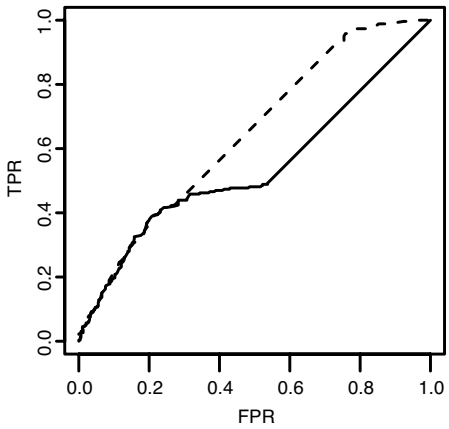


(a) ROC curve of gtt.

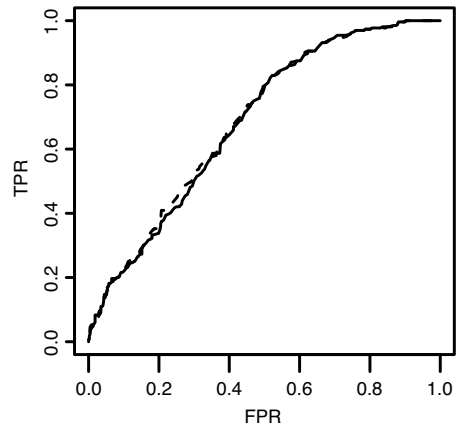
(b) ROC curve of dbp.



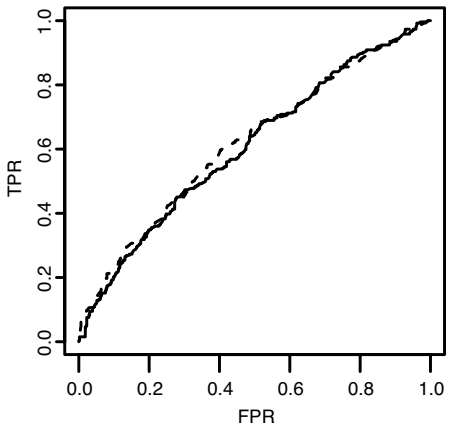
(c) ROC curve of tsft.



(d) ROC curve of si.



(e) ROC curve of bmi.



(f) ROC curve of dpf.

Fig. 2. The estimated ROC curves of individual biomarkers (solid line) and their optimal scoring functions (dashed line).

original measurements, these biomarkers are still not acceptable under the rule of thumb suggested by [10]. To improve the poor classification of blood tests, our next strategy is to seek an optimal combination of these blood tests. By applying the proposed test rules, no significant difference is detected for the ROC curves of the composite biomarkers $g(mcv, alphos, sgpt, sgot, \gamma gt)$ and $g_3^5(sgpt, sgot, \gamma gt)$ ($p = 0.990$; $p^* = 0.979$). The AUC estimate 0.783 (0.0242; 0.0278) of $g_3^5(sgpt, sgot, \gamma gt)$ indicates a near excellent discriminating ability. Compared with the AUCs of γgt and an optimal transformation of *sgpt*, about 24%–25% increments are obtained in the derived optimal composite biomarker.

Table 10
The estimates for AUCs of five biomarkers and their optimal transformations, the standard errors (se and se^*) and the p-values (p and p^*) based on the normal and bootstrap testing procedures.

Biomarker	$\widehat{\theta}_{Y_s}$	$se(\widehat{\theta}_{Y_s})$	$se^*(\widehat{\theta}_{Y_s})$	$\widehat{\theta}_{\widehat{g}_1^5}$	$se(\widehat{\theta}_{\widehat{g}_1^5})$	$se^*(\widehat{\theta}_{\widehat{g}_1^5})$	p	p^*
mcv	0.560	0.0307	0.0310	0.621	0.0301	0.0302	0.026	0.056
$alphos$	0.571	0.0310	0.0311	0.571	0.0308	0.0376	0.506	0.763
$sgpt$	0.579	0.0306	0.0307	0.634	0.0298	0.0306	0.010	0.072
$sgot$	0.584	0.0310	0.0311	0.613	0.0303	0.0350	0.144	0.314
γgt	0.628	0.0307	0.0199	0.629	0.0299	0.0348	0.484	0.824

The estimate 0.790 (0.0246; 0.0247) of $\theta_{g_{3si}^5}$ is obtained via using the estimated predictor $\widehat{g}_{3si}^5(sgpt - 0.420sgot - 0.617\gamma gt)$ and there is no strong evidence to reject the optimality of $g_{3si}^5(sgpt + \beta_1 sgot + \beta_2\gamma gt)$ ($p = 0.617, p^* = 0.762$). The parametric and semiparametric models in the first data analysis were continuously applied and compared. By using the estimated linear predictors $Y_{\widehat{\beta}_{slogit}} = 0.933 + 0.056sgpt - 0.103sgot - 0.013\gamma gt$ and $Y_{\widehat{\beta}_{sprobit}} = 0.546 + 0.032sgpt - 0.061sgot - 0.007\gamma gt$, the estimates 0.707 (0.0290; 0.0296) of $\theta_{Y_{\beta_{logit}}}$ and 0.706 (0.0290; 0.0298) of $\theta_{Y_{\beta_{sprobit}}}$ indicate that $Y_{s\beta_{logit}}$ and $Y_{\beta_{sprobit}}$ have only acceptable predictive abilities. It is further concluded from the p-values (0.003; 0.001) for testing $\mathcal{R}_{g_{3si}^5}(u) = \mathcal{R}_{Y_{\beta_{slogit}}}(u)$ and the p-values (0.003; 0.002) for testing $\mathcal{R}_{g_{3si}^5}(u) = \mathcal{R}_{Y_{\beta_{sprobit}}}(u)$ that these link functions might be misspecified.

6. Concluding remarks

In this study, we proposed nonparametric estimators for the accuracy measures of optimal transformations of individual markers and optimal composite markers derived from a semiparametric or nonparametric model. The form of hypotheses can be used to check whether a marker itself is optimal, to explore important multiple markers, and investigate the adequacy of single-index models. The difference of proposed AUC estimators is also adopted as a test statistic in the establishment of test rules. In our extensive numerical study, the proposed methods were demonstrated to have satisfactory performances.

A major concern in tackling with the above issues is that the conventional approaches for comparing the classification capabilities of two diagnostic systems or test results might be problematic. As stressed in our theoretical justification, the asymptotic results of $\widehat{\theta}_{g_m^q}$ cannot be generalized to $\widehat{\theta}_{\widehat{g}_m^q}$ when $g_m^q(y_s)$ is flat in some regions of positive probability. A more thorough investigation of this circumstance would be desirable and is planned for future studies. For optimal transformations of individual markers, their ROC curves might cross each other and have the same or similar AUCs. As expected, a test based on the difference of estimated AUCs should have a poor statistical power and the permutation tests (cf. [3,24,23]) are recommended. The validity of such inference procedures can be verified by applying the arguments in the proofs of Theorems 1–2.

It has been shown in our numerical experiments that a more accurate classification usually associates with the correct model. Although a linear combination of markers is preferred by practitioners, more complex nonlinear models are also typically required in some biological fields. As we can see, the nonparametric smoothing approach is widely used for exploratory data analysis. One of our important contributions is that the proposed test statistic $\widehat{\theta}_{\widehat{g}_{msi}^q} - \widehat{\theta}_{Y_{\beta_s}}$ can be used to check the monotonicity of $g_{msi}^q(y_{\beta_s})$ in y_{β_s} and the correctness of a fully parametric model specification. Another competing approach to improve the power of tests is to adopt the lack-of-fit model checking of [25] with some modifications. When the dimension of markers is not low, it becomes a challenging issue to discover significant markers. Our future study will aim to develop an efficient procedure which simultaneously identifies the true model and achieves optimal classification.

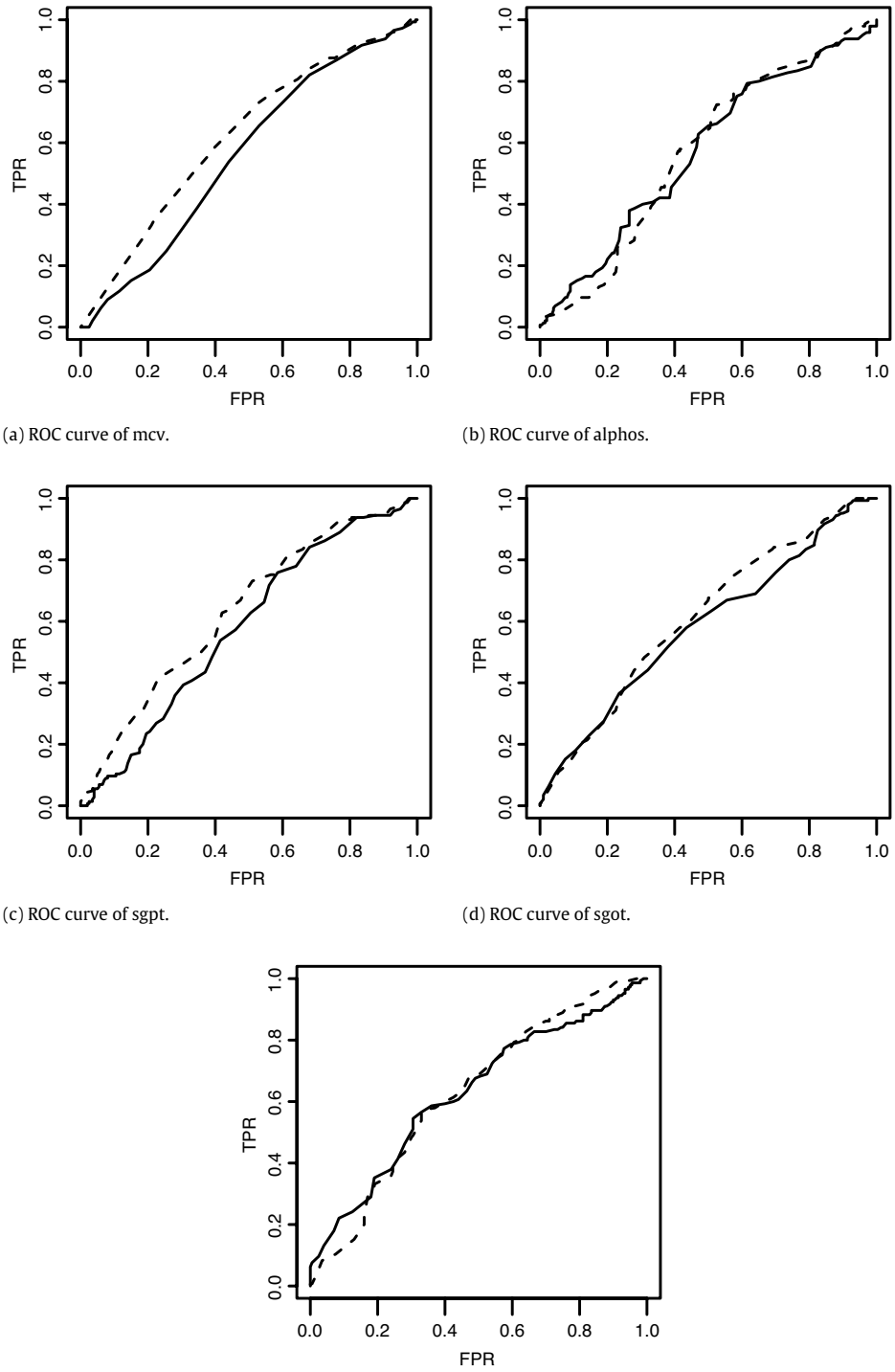
In the framework of survival data, a recent extension of the ROC curve analysis has focused on the binary time-varying classification task. Continuous markers are frequently measured at or before the outset of study and the binary outcome is defined over time based on the time of a specific disease event or the death time T. With a scalar-valued function $g_{mt}^q(Y_s)$, [8] generalized the ROC curve as the plot of the time-dependent true positive rate $P(g_{mt}^q(Y_s) > c|T \le t)$ versus the time-dependent false positive rate $P(g_{mt}^q(Y_s) > c|T > t)$ for various values of threshold c. It is easy to show that the conditional survival function $S(t|Y)$ is an optimal composite marker with the highest ROC curve at each time point t. Our main results should be valuable in the development of inference for its accuracy measure.

Acknowledgments

The research of the authors was partially supported by the National Science Council grants 97-2118-M-002-020-MY2 and 99-2118-M-002-003 (Taiwan). We would like to thank the associate editor and two reviewers for valuable comments on the article.

Appendix

Proof of Theorem 1. In the following proof, we will show that $E[n(\widehat{\theta}_{\widehat{g}_m^q} - \widehat{\theta}_{g_m^q})^2] = o(1)$ because Theorem 1 is a direct consequence of this property. Let $q_{ijs} = \{I(\widehat{g}_{ijs} > 0) - I(g_{ijs} > 0)\} + 0.5\{I(\widehat{g}_{ijs} = 0) - I(g_{ijs} = 0)\}$ with $g_{ijs} = g_m^q(Y_{is}) - g_m^q(Y_{js})$



(a) ROC curve of mcv.

(b) ROC curve of alphos.

(c) ROC curve of sgpt.

(d) ROC curve of sgot.

(e) ROC curve of gamma-gt.

Fig. 3. The estimated ROC curves of individual biomarkers (solid line) and their optimal scoring functions (dashed line).

and $\widehat{g}_{ijs} = \widehat{g}^q_m(Y_{is}) - \widehat{g}^q_m(Y_{js}), i, j = 1, \ldots, n.\, \widehat{\theta}_{\widehat{g}^q_m} - \widehat{\theta}_{g^q_m}$ can be re-expressed as

$$\widehat{\theta}_{\widehat{g}^q_m} - \widehat{\theta}_{g^q_m} = \frac{\sum_{i \neq j} q_{ijs} D_{ij}}{n(n-1)p_0 p_1}(1 + o_p(1)). \tag{A.1}$$

It follows that

$$E[n(\widehat{\theta}_{g_m^q} - \widehat{\theta}_{g_m^q})^2] = \left\{ n^{-1}E[q_{12s}D_{12}] + E[q_{12s}D_{12}(q_{13s}(1-D_3) + q_{32s}D_3)] \right.$$
$$\left. + \; nE\left[\prod_{\ell=1,3} q_{\ell(\ell+1)s}D_{\ell(\ell+1)}\right]\right\}(p_0p_1)^{-2}(1+o(1)). \tag{A.2}$$

Since $|q_{12s}D_{12}| \le 1$, the first term on the right-hand side of (A.2) is $O(n^{-1})$. By the continuous mapping theorem, q_{12s} is ensured to converge in probability to zero for (y_{1s}, y_{2s}) satisfying $g_{12s} \neq 0$. The bounded convergence theorem together with assumption (A4) imply that the second term is $o(1)$. Thus, one has

$$E[n(\widehat{\theta}_{g_m^q} - \widehat{\theta}_{g_m^q})^2] = nE\left[\prod_{\ell=1,3} q_{\ell(\ell+1)s}D_{\ell(\ell+1)s}\right](p_0p_1)^{-2}(1+o(1)). \tag{A.3}$$

Let $q_{12s}^{-(i-\ell)}$ be defined as q_{12s} with the individuals indexed by i to ℓ being deleted. We can derive that

$$\left|E\left[\prod_{\ell=1,3} q_{\ell(\ell+1)s}^{-(1-4)}D_{\ell(\ell+1)}\right]\right| = \frac{1}{4}\left|E\left[\prod_{\ell=1,3}\int q_{\ell(\ell+1)s}^{-(1-4)}g_{\ell(\ell+1)s}f(y_{\ell s})f(y_{(\ell+1)s})\mathrm{d}y_{\ell s}\mathrm{d}y_{(\ell+1)s}\right]\right|$$
$$\le \frac{1}{4}\left\{\int (E[|q_{12s}^{-(1-4)}|])^{1/2}|g_{12s}|f(y_{1s})f(y_{2s})\mathrm{d}y_{1s}\mathrm{d}y_{2s}\right\}^2, \tag{A.4}$$

where $E[|q_{12s}^{-(1-4)}|] = P(|q_{12s}^{-(1-4)}| = 1, M_{12s} > |g_{12s}|/2) + P(|q_{12s}^{-(1-4)}| = 1, M_{12s} \le |g_{12s}|/2)$ and $M_{12s} = \max\{|\widehat{g}_m^q(y_{1s}) - g_m^q(y_{1s})|, |\widehat{g}_m^q(y_{2s}) - g_m^q(y_{2s})|\}$. Since $M_{12s} \le |g_{12s}|/2$, the property $q_{ij}^{-(i-j)} = 0$ is directly obtained. Together with the large deviation probability for $\widehat{g}_m^q(y_s)$ (cf. [14,16]), it yields that

$$E[|q_{12s}^{-(1-4)}|] = P\left(|q_{12s}^{-(1-4)}| = 1, M_{12s} > \frac{|g_{12s}|}{2}\right) \le P\left(M_{12s} > \frac{|g_{12s}|}{2}\right) \le \mathrm{e}^{-nh(\Lambda_{12s}(\frac{|g_{12s}|}{2})+\eta_{12s})}, \tag{A.5}$$

where $\Lambda_{12s}(u) = \min\{\gamma_{y_{1s}}(u), \gamma_{y_{2s}}(u)\}$ and $\eta_{12s} = o(|g_{12s}|/2)$ with $\lim_{u\to 0}\sup_{y_s}|\gamma_{y_s}(u)/\gamma_{y_s}^*(u)| = 1$ and $\gamma_{y_s}^*(u) = u^2f(y_s)[2g_m^q(y_s)/\{1-g_m^q(y_s)\}]\int K^2(v)\mathrm{d}v$. Combining (A.4) and (A.5), the following inequality is ascertained:

$$E\left[\prod_{\ell=1,3} q_{\ell(\ell+1)s}^{-(14)}D_{\ell(\ell+1)}\right] \le \left(\int \mathrm{e}^{-\frac{nh}{2}\Lambda_{12s}(\frac{|g_{12s}|}{2})}|g_{12s}|f(y_{1s})f(y_{2s})\mathrm{d}y_{1s}\mathrm{d}y_{2s}\right)^2(1+o(1)). \tag{A.6}$$

For any $\delta > 0$, there exists $u \in (0,\delta)$ satisfying $\gamma_{y_s}^*(u)/2 \le \gamma_{y_s}(u) \le 3\gamma_{y_s}^*(u)/2$. Let $\Lambda_{12s}^*(u) = \min\{\gamma_{y_{1s}}^*(u), \gamma_{y_{2s}}^*(u)\}$, $R_1 = \{(y_{1s}, y_{2s}) : \Lambda_{12s}(|g_{12s}|/2) \ge \Lambda_{12s}^*(|g_{12s}|/2)\}$, $R_2 = \{(y_{1s}, y_{2s}) : \Lambda_{12s}(|g_{12s}|/2) < \Lambda_{12s}^*(|g_{12s}|/2), |g_{12s}|/2 \le \delta\}$, $R_3 = \{(y_{1s}, y_{2s}) : \Lambda_{12s}(|g_{12s}|/2) < \Lambda_{12s}^*(|g_{12s}|/2), |g_{12s}|/2 > \delta\}$, and $\mathbb{R}_l = \int_{R_l}\mathrm{e}^{-nh/2}\Lambda_{12s}(|g_{12s}|/2)|g_{12s}|f(y_{1s})f(y_{2s})\mathrm{d}y_{1s}\mathrm{d}y_{2s}$, $l = 1, 2, 3$. By direct calculations and assumption (A4), it is easy to show that

$$\mathbb{R}_1 \le \int \mathrm{e}^{-\frac{nh}{2}\Lambda_{12s}^*(\frac{|g_{12s}|}{2})}|g_{12s}|f(y_{1s})f(y_{2s})\mathrm{d}y_{1s}\mathrm{d}y_{2s} \le \sum_{k=0}^{\lceil nh\rceil}\int_{\{\frac{k}{nh}\le\Lambda_{12s}^*(\frac{|g_{12s}|}{2})\le\frac{k+1}{nh}\}}\mathrm{e}^{-\frac{nh}{4}\Lambda_{12s}^*(\frac{|g_{12s}|}{2})}$$
$$|g_{12s}|f(y_{1s})f(y_{2s})\mathrm{d}y_{1s}\mathrm{d}y_{2s} \le \sum_{k=0}^{\lceil nh\rceil}\mathrm{e}^{-k}\left(\frac{k+1}{nh}\right)^{1/2}M_f^{1/2}I\left(\frac{4\int K^2(v)\mathrm{d}v}{nh}\right) = o(n^{-1/2}) \tag{A.7}$$

$$\text{and } \mathbb{R}_2 \le \int \mathrm{e}^{-\frac{nh}{4}\Lambda_{12s}^*(\frac{|g_{12s}|}{2})}|g_{12s}|f(y_{1s})f(y_{2s})\mathrm{d}y_{1s}\mathrm{d}y_{2s} = o(n^{-1/2}). \tag{A.8}$$

Since there exists $d > 0$ satisfying $\Lambda_{y_{1s}y_{2s}}(\delta) \ge \Lambda_{y_{1s}y_{2s}}^*(\delta)/d$, $\mathbb{R}_3 = o(n^{-1/2})$ is ensured from assumption (A4) and the inequality

$$\mathbb{R}_3 \le \int \mathrm{e}^{-\frac{nh}{2}\Lambda_{12s}(\delta)}|g_{12s}|f(y_1)f(y_{2s})\mathrm{d}y_{1s}\mathrm{d}y_{2s} \le 2\delta\int \mathrm{e}^{-\frac{nhfm(y_{1s},y_{2s})}{2d\int K^2(v)\mathrm{d}v}}f(y_1)f(y_{2s})\mathrm{d}y_{1s}\mathrm{d}y_{2s}. \tag{A.9}$$

From (A.7)–(A.9), the term $nE[\prod_{\ell=1,3} q_{-(1-4)}(Y_{\ell s}, Y_{\ell+1s})D_\ell(1-D_{\ell+1})]$ in (A.6) is shown to be

$$nE\left[\prod_{\ell=1,3} q_{\ell(\ell+1)s}^{-(1-4)}D_{\ell(\ell+1)}\right] = o(1). \tag{A.10}$$

Along much the same lines as the arguments in (A.4)–(A.6), one can also obtain that

$$q_{\ell(\ell+1)s} = q^{-(1-4)}_{\ell(\ell+1)s}(1 + o_p(1)), \quad \ell = 1, 3. \tag{A.11}$$

Finally, the property $E[n(\widehat{\theta}_{g^q_m} - \widehat{\theta}_{g^q_m})^2] = o(1)$ is implied from (A.3), (A.10) and (A.11). □

Proof of Theorem 2. The first-order Taylor expansion of $\widehat{g}^q_{msi}(y_{\widehat{\beta}_s})$ at $\widehat{\beta}_s = \beta_s$ leads to

$$\widehat{g}^q_{msi}(y_{\widehat{\beta}_s}) = \widehat{g}^q_{msi}(y_{\beta_s}) + \left(\frac{\partial \widehat{g}^q_{msi}(y_{\bar{\beta}_s})}{\partial \beta_s}\right)^{\top} (\widehat{\beta}_s - \beta_s), \tag{A.12}$$

where $\bar{\beta}_s$ lies on the line segment between $\widehat{\beta}_s$ and β_s. It follows from Lemmas 5.1 and 5.7–5.10 of [11] that

$$|\widehat{g}^q_{msi}(y_{\widehat{\beta}_s}) - g^q_{msi}(y_{\beta_s})| \le |\widehat{g}^q_{msi}(y_\beta) - g_{sim}(y_\beta)| + \frac{M_0}{n}\sum_{i=1}^{n} |D_i - g^q_{msi}(Y_{i\beta_s})| + \gamma(y_s, \beta_s) \tag{A.13}$$

for some positive constant M_0 and $\sup_{y_s \times \mathcal{B}_s} \gamma(y_s, \beta) = o_p(n^{-1/2})$. Let $M_{12\beta_s} = \max\{|\widehat{g}^q_{msi}(y_{1\beta_s}) - g^q_{msi}(y_{1\beta_s})|, |\widehat{g}^q_{msi}(y_{2\beta_s}) - g^q_{msi}(y_{2\beta_s})|\}$ and $g_{12\beta_s} = g^q_{msi}(y_{1\beta_s}) - g^q_{msi}(y_{2\beta_s})$. Similar calculations to those leading to (A.5) yield that

$$P\left(M_{12\beta_s} > \frac{|g_{12\beta_s}|}{6}\right) \le \mathrm{e}^{\frac{-nh(g_{12\beta_s})^2 m_{f_{\beta_s}}}{18} + \eta_{12\beta_s}}, \tag{A.14}$$

where $\eta_{12\beta_s} = o(|g_{12\beta_s}|/2)$. By Bernstein's inequality and Chebyshev's inequality, one has

$$P\left(\frac{M_0}{n}\sum_{i=1}^{n} |D_i - g^q_{msi}(Y_{i\beta_s})| > \frac{|g_{12\beta_s}|}{6}\right) \le \mathrm{e}^{\frac{-ng^2_{12\beta_s}}{M_0(18M_0+16)}} \quad \text{and} \quad P\left(\gamma(y_s, \beta_s) > \frac{|g_{12\beta_s}|}{6}\right) = o(n^{-1}). \tag{A.15}$$

Together with (A.14), Theorem 2 is established paralleling the proof of Theorem 1.

Proof of Theorem 4. Using the bootstrap variance computation of [6], one has

$$V^*(\widehat{\theta}^*_g) = \frac{(n_1 - 1)(\alpha_0 - \widehat{\theta}^2_g) + (n_0 - 1)(\alpha_1 - \widehat{\theta}^2_g) + \widehat{\theta}_g(1 - \widehat{\theta}_g)}{n_0 n_1}, \tag{A.16}$$

where

$$\alpha_\ell = n_\ell^{-1}\sum_{i=1}^{n}\left\{n_{1-\ell}^{-1}\sum_{j=1}^{n}\eta_{ji}D^\ell_{1-\ell}(1 - D_j)^\ell\right\}^2 D^\ell_i(1 - D_i)^{1-\ell}, \quad \ell = 0, 1.$$

Straightforward calculations ensure that

$$E[|\widehat{\alpha}_0 - \alpha_0|] \le E\left[\left|\left(n_1^{-1}\sum_{j=1}^{n}\widehat{\eta}_{ji}D_j\right)^2 - \left(n_1^{-1}\sum_{j=1}^{n}\eta_{ji}D_j\right)^2\right|\right] \le E[|\widehat{\eta}_{ji}\widehat{\eta}_{ki} - \eta_{ji}\eta_{ki}|]. \tag{A.17}$$

Since $E[|\widehat{\eta}_{ji}\widehat{\eta}_{ki} - \eta_{ji}\eta_{ki}|]$ is bounded above by one and converges to zero in probability, it follows immediately that $E|\widehat{\alpha}_0 - \alpha_0| = o(1)$ and, hence, $\widehat{\alpha}_0 - \alpha_0 \xrightarrow{p} 0$. Similarly, we can derive that $\widehat{\alpha}_1 - \alpha_1 \xrightarrow{p} 0$. Together with $\sqrt{n}(\widehat{\theta}_{\widehat{g}} - \widehat{\theta}_g) \xrightarrow{p} 0$ in Theorems 1–2 and $nV^*(\widehat{\theta}^*_g) - \sigma^2_g \xrightarrow{p} 0$, $n(V^*(\widehat{\theta}^*_{\widehat{g}}) - V^*(\widehat{\theta}^*_g)) \xrightarrow{p} 0$ and $nV^*(\widehat{\theta}^*_{\widehat{g}}) - \sigma^2_g \xrightarrow{p} 0$ are ascertained. By further applying Slutsky's theorem, (3.6) is established.

By using $\widehat{\theta}^*_{\widehat{g}} - \widehat{\theta}_{\widehat{g}} = (\widehat{\theta}^*_{\widehat{g}} - \widehat{\theta}^*_g) + (\widehat{\theta}^*_g - \widehat{\theta}_g) + (\widehat{\theta}_g - \widehat{\theta}_{\widehat{g}})$, the proof of (3.7) can be developed by verifying

$$P^*(\sqrt{n}(\widehat{\theta}^*_g - \widehat{\theta}_g) \le z) - P(\sqrt{n}(\widehat{\theta}_{\widehat{g}} - \theta_g) \le z) \xrightarrow{p} 0 \quad \text{and} \tag{A.18}$$

$$P^*(\sqrt{n}|(\widehat{\theta}^*_{\widehat{g}} - \widehat{\theta}^*_g) + (\widehat{\theta}_g - \widehat{\theta}_{\widehat{g}})| > \varepsilon) \xrightarrow{p} 0 \quad \text{for any } \varepsilon > 0. \tag{A.19}$$

Combining the result of [2] and Theorems 1–2, the property in (A.18) holds. It is implied from (3.4), (A.16), and the Chebyshev's inequality that

$$P^*(\sqrt{n}|(\widehat{\theta}^*_{\widehat{g}} - \widehat{\theta}^*_g) + (\widehat{\theta}_g - \widehat{\theta}_{\widehat{g}})| > \varepsilon) \le \frac{n\left\{\sum_{l=0}^{1}(n_{1-\ell} - 1)(\alpha_\ell + \widehat{\alpha}_\ell - 2\kappa_\ell) + \kappa_2\right\}}{n_0 n_1 \varepsilon^2}, \tag{A.20}$$

where $\kappa_\ell = n_\ell^{-1}\sum_{i=1}^n\{n_{1-\ell}^{-1}\sum_{j=1}^n \eta_{ij}D_j^{1-\ell}(1-D_j)^\ell\}\{n_1^{-1}\sum_{j=1}^n \widehat{\eta}_{ij}D_j^{1-l}(1-D_j)^\ell\}D_i^l(1-D_i)^{1-\ell}$, $\ell = 0, 1$, and $\kappa_2 = \widehat{\theta}_g(1-\widehat{\theta}_g) + \widehat{\theta}_{\widehat{g}}(1-\widehat{\theta}_{\widehat{g}}) + 2\widehat{\theta}_g\widehat{\theta}_{\widehat{g}} - 2\sum\sum_{i\neq j}\eta_{ij}\widehat{\eta}_{ij}D_i(1-D_j)/(n_0n_1)$. Paralleling the proof of (A.17), it follows that $\alpha_\ell + \widehat{\alpha}_\ell - 2\kappa_\ell \xrightarrow{p} 0$, $\ell = 0, 1$. Together with $\kappa_2 = O(1)$, $nV^*(\widehat{\theta}_{\widehat{g}}^* - \widehat{\theta}_g^*)/\varepsilon^2$ in (A.20) is shown to be $o_p(1)$ and, hence, (A.19) is obtained. □

References

[1] R. Bellman, Adaptive Control Processes: A Guide Tour, New Jersey, Princeton, 1961.
[2] P.J. Bickel, D.A. Freedman, Some asymptotic theory for the bootstrap, Ann. Statist. 9 (1981) 1196–1217.
[3] T.M. Braun, T.A. Alonzo, A modified sign test for comparing paired ROC curves, Biostatistics 9 (2008) 364–372.
[4] R.D. Cook, H. Lee, Dimension reduction in binary response regression, J. Amer. Statist. Assoc. 94 (1999) 1187–1201.
[5] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the area under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (1988) 837–845.
[6] B. Efron, Bootstrap methods: another look at the Jackknife, Ann. Statist. 7 (1979) 1–26.
[7] W. Härdle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, Ann. Statist. 21 (1993) 157–178.
[8] P.J. Heagerty, T. Lumley, M.S. Pepe, Time-dependent ROC curves for censored survival data and a diagnostic marker, Biometrics 54 (2000) 124–135.
[9] W. Hoeffding, A class, of statistics with asymptotically normal distribution, Ann. Statist. 19 (1948) 293–325.
[10] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Wiley, New York, 2000.
[11] H. Ichimura, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, J. Econom. 58 (1993) 71–120.
[12] R.W. Klein, R.H. Spady, An efficient semiparametric estimator for binary response models, Econometrica 61 (1993) 387–421.
[13] S. Ma, J. Huang, Combining multiple markers for classification using ROC, Biometrics 63 (2007) 751–757.
[14] S.M.O. Maouloud, Some uniform large deviation results in nonparametric function estimation, J. Nonparametr. Stat. 20 (2008) 129–152.
[15] M.S. McIntosh, M.S. Pepe, Combining several screening tests: optimality of the risk score, Biometrics 58 (2002) 657–664.
[16] A. Mokkadem, M. Pelletier, B. Thiam, Large and moderate deviations principles for kernel estimators of a multivariate regression, Math. Methods Statist. 17 (2008) 146–172.
[17] M.S. Pepe, T. Cai, G. Longton, Combining predictors for classification using the area under the receiver operating characteristic curve, Biometrics 62 (2006) 221–229.
[18] G. Qin, X.H. Zhou, Empirical likelihood inference for the area under the ROC curve, Biometrics 62 (2006) 613–622.
[19] A.J. Scott, C.J. Wild, Fitting regression models to case-control data by maximum likelihood, Biometrika 84 (1997) 57–71.
[20] S.J. Sheather, M.C. Jones, A reliable data-based bandwidth selection method for kernel density estimation, J. R. Stat. Soc. B53 (1991) 683–690.
[21] D.F. Signorini, M.C. Jones, Kernel estimators for univariate binary regression, J. Amer. Statist. Assoc. 99 (2004) 119–126.
[22] J.W. Smith, J.E. Everhart, W.C. Dickson, W.C. Knowler, R.S. Johannes, Using the adap learning algorithm to forecast the onset of diabetes mellitus, in: Proceedings of the Symposium on Computer Applications and Medical Care, IEEE Computer Society Press, 1988, pp. 261–265.
[23] E.S. Venkatraman, A permutation test to compare receiver operating characteristic curves, Biometrics 56 (2000) 1134–1138.
[24] E.S. Venkatraman, C.B. Begg, A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment, Biometrika 83 (1996) 835–848.
[25] Y. Xia, Model checking in regression via dimension reduction, Biometrika 96 (2009) 133–148.
[26] M. Yalçın, T. Yıldırım, Karaciğerbozukluklarının yapay sinir ağlarıile teşhisi, in: biyomedikal mühendisliği ulusal toplantısı, Istanbul, Türkiye, 2003, pp. 293–297.